3952-37

McB

STRATEGIES FOR MEDICAL TECHNOLOGY ASSESSMENT

SEPTEMBER 1982

vii+227 p. il. Index. Oversized.

236 p.

LC. 82-600606.
SN 052-003-00887-4.

CONGRESS OF THE UNITED STATES
Office of Technology Assessment
Washington, D.C. 20510

9/81

J952-37

A

Office of Technology Assessment

Congressional Board of the 97th Congress

TED STEVENS, *Alaska, Chairman*

MORRIS K. UDALL, *Arizona, Vice Chairman*

Senate

ORRIN G. HATCH
Utah

CHARLES McC. MATHIAS, JR.
Maryland

EDWARD M. KENNEDY
Massachusetts

ERNEST F. HOLLINGS
South Carolina

HOWARD W. CANNON
Nevada

House

GEORGE E. BROWN, JR.
California

JOHN D. DINGELL
Michigan

LARRY WINN, JR.
Kansas

CLARENCE E. MILLER
Ohio

COOPER EVANS
Iowa

JOHN H. GIBBONS
(Nonvoting)

Advisory Council

CHARLES N. KIMBALL, *Chairman*
Midwest Research Institute

EARL BEISTLINE
University of Alaska

CHARLES A. BOWSHER
General Accounting Office

CLAIRE T. DEDRICK
California Land Commission

JAMES C. FLETCHER
University of Pittsburgh

S. DAVID FREEMAN
Tennessee Valley Authority

GILBERT GUDE
Congressional Research Service

CARL N. HODGES
University of Arizona

RACHEL McCULLOCH
University of Wisconsin

WILLIAM J. PERRY
Hambrecht & Quist

DAVID S. POTTER
General Motors Corp.

LEWIS THOMAS
Memorial Sloan-Kettering
Cancer Center

Director

JOHN H. GIBBONS

The Technology Assessment Board approves the release of this report. The views expressed in this report are not necessarily those of the Board, OTA Advisory Council, or of individual members thereof.

STRATEGIES FOR MEDICAL TECHNOLOGY ASSESSMENT

SEPTEMBER 1982

OTA Reports are the principal documentation of formal assessment projects. These projects are approved in advance by the Technology Assessment Board. At the conclusion of a project, the Board has the opportunity to review the report but its release does not necessarily imply endorsement of the results by the Board or its individual members.

CONGRESS OF THE UNITED STATES
Office of Technology Assessment
Washington, D.C. 20510

Library of Congress Catalog Card Number 82-600606

For sale by the Superintendent of Documents,
U.S. Government Printing Office, Washington, D.C. 20402

Foreword

This report, *Strategies for Medical Technology Assessment,* analyzes the present system of identifying and testing medical technologies and of synthesizing and disseminating assessment information. OTA began the study in July 1980, at the request of the House Committee on Energy and Commerce.

The report focuses on the flow of information that is central to an efficient assessment system. Methods for testing technologies and for synthesizing information are explored, and a compendium of data and bibliographic sources are included. The report also describes the innovation process for medical technologies, the effects that Federal policies have on that process, and the needs those policies generate for technology assessment information. It critiques the current system of assessment and provides policy options, both legislative and oversight, for Congress to improve the system.

During the course of this assessment, both the House Committee on Energy and Commerce and the Senate Committee on Labor and Human Resources requested that OTA study several specific areas in more depth. In response to these requests, OTA is publishing three other volumes: 1) a report on medical technology under proposals to increase competition in health care, 2) a report on the postmarketing surveillance of prescription drugs, and 3) a technical memorandum on MEDLARS (the National Library of Medicine's Medical Literature Analysis and Retrieval System) and health information policy. Another paper, funded as part of this assessment, concerns the potential role of Professional Standards Review Organizations in medical technology assessment.

In preparing this report, OTA consulted with members of the advisory panel for the assessment, with contractors and special consultants, and with numerous other experts in industry, medicine, economics, pharmacology, ethics, information science, and health policy.

Drafts of the final report were reviewed by the advisory panel chaired by Dr. Lester Breslow, by the Health Program Advisory Committee chaired by Dr. Sidney S. Lee, and by approximately 100 other individuals and groups representing a wide range of disciplines and perspectives. We are grateful for their many contributions. As with all OTA reports, however, the content is the responsibility of the Office and does not constitute consensus or endorsement by the advisory panel or the Technology Assessment Board.

John H. Gibbons

JOHN H. GIBBONS
Director

Advisory Panel for Strategies for Medical Technology Assessment

Lester Breslow, *Panel Chairman*
School of Public Health, University of California, Los Angeles

Morris Collen
Director of Technology Assessment
Kaiser Permanente Medical Group

Richard Cooper
Williams & Connolly, Inc.

D. V. d'Arbeloff
Chairman and Chief Executive Officer
Millipore Corp.

Harvey Fineberg
Harvard School of Public Health

Jerome D. Frank
The Henry Phipps Psychiatric Clinic
The Johns Hopkins Hospital

William Goffman
School of Library Science
Case Western Reserve University

Leon Greene
Vice President
New Product Technology
Smith, Kline & French Laboratories, Inc.

David B. Horner
Orthopedic Surgeon

Stanley B. Jones
Vice President
Blue Cross/Blue Shield Association

F. Wilfrid Lancaster
Graduate School of Library and Information Science
University of Illinois

Louise B. Russell
Senior Fellow
The Brookings Institute

Herbert Semmel
President
Consumer Coalition for Health

Robert M. Veatch
Kennedy Institute of Ethics
Georgetown University

Richard W. Vilter
American College of Physicians
College of Medicine
University of Cincinnati

Kenneth E. Warner
School of Public Health
University of Michigan

Richard N. Watkins
Staff Physician
Group Health Cooperative

Carol Weiss
Graduate School of Education
Harvard University

Kerr L. White
Deputy Director for Health Sciences
Rockefeller Foundation

OTA Project Staff—Strategies for Medical Technology Assessment

H. David Banta, *Assistant Director, OTA*
Health and Life Sciences Division

Clyde J. Behney, *Health Program Manager*

Bryan R. Luce, *Project Director*
Dale A. Carlson, *Analyst*
John C. Langenbrunner, *Analyst*
Gloria Ruby, *Analyst*
Kerry Britten Kemp, *Editor*

Virginia Cwalina, *Administrative Assistant*
Mary E. Harvey, *Secretary*
Pamela Simerly, *Secretary*
Lorraine G. Ferris, *Secretary**
Nancy L. Kenney, *Secretary**

Other Contributing Staff

Ann Rose, *Senior Analyst*
Arthur Kohrman, *Congressional Fellow*
Hellen Gelband, *Analyst*
Mark Hochler, *Intern*
Lisa Scheffler, *Intern*
Harvey Sepler, *Intern*

Principal Contractors

Rand Corp.
John P. Bunker, *Stanford University*
Lawrence Miike
John B. Reiss, *Baker and Hostetler*
Leonard Saxe, *Boston University*
Paul M. Wortman, *University of Michigan*

OTA Publishing Staff

John C. Holmes, *Publishing Officer*

John Bergling Kathie S. Boss Debra M. Datcher Joe Henson

---

*Until January 1982.
**Until September 1981.

Health Program Advisory Committee

Sidney S. Lee, *Chairman*
Vice President, Michael Reese Hospital and Medical Center

Stuart H. Altman
Florence Heller School
Brandeis University

Robert Ball*
Institute of Medicine
National Academy of Sciences

Lewis N. Butler*
Health Policy Program
School of Medicine
University of California, San Francisco

Kurt Deuschle
Mount Sinai School of Medicine

Zita Fearon*
Consumer Commission on the Accreditation of Health Services, Inc.

Rashi Fein
Center for Community Health and Medical Care
Harvard Medical School

Melvin A. Glasser
Committee for National Health Insurance

Patricia King
Georgetown Law Center

Joyce C. Lashof
School of Public Health
University of California, Berkeley

Mark Lepper
Vice President for Inter-Institutional Affairs
Rush-Presbyterian-St. Luke's Medical Center

Margaret Mahoney
President
The Commonwealth Fund

Frederick Mosteller
Department of Health Policy and Management
Harvard School of Public Health

Beverlee Myers*
Director
Department of Health Services
State of California

Mitchell Rabkin
General Director
Beth Israel Hospital

Dorothy P. Rice
Department of Social and Behavioral Sciences
School of Nursing
University of California, San Francisco

Richard K. Riegelman
School of Medicine
George Washington University

Walter L. Robb
Vice President and General Manager
Medical Systems Operations
General Electric

Frederick C. Robbins
President
Institute of Medicine
National Academy of Sciences

Rosemary Stevens
Department of History and Sociology of Science
University of Pennsylvania

Kerr White
Deputy Director for Health Sciences
Rockefeller Foundation

---

*Former members of the Health Program Advisory Committee whose terms expired prior to the completion of this study.

Contents

Page

1. Introduction and Summary ... 3
2. Information for Assessment ... 23
3. Evaluating Health and Economic Effects ... 33
4. Social Values in Technology Assessment ... 45
5. Synthesis, Cost-Effectiveness Analysis, and Decisionmaking ... 57
6. Factors Affecting the Development, Diffusion, and Use of Medical Technologies ... 71
7. Critique of the Current System ... 91
8. Policy Options ... 105

Appendixes

A. Compendium of Statistical Data Sources for Medical Technology Assessment ... 114
B. Compendium of Bibliographic Data Bases for Medical Technology Assessment ... 121
C. Assessment of Medical Technology: Methodological Considerations ... 127
D. Medical Technologies and Innovation ... 150
E. Case Studies of Medical Innovations ... 167
F. Model for an Institute for Health Care Evaluation ... 185
G. Method of the Study and Description of Other Volumes ... 192
H. Acknowledgments ... 195
I. Glossary of Acronyms and Terms ... 198

References ... 205

Index ... 223

1.
Introduction and Summary

Knowledge advances by steps, and not by leaps.

—*Thomas Babbington Macaulay*

Contents

Page

Need for a Strategy 3
Dimensions of the Need and the Problem 4
  Government 4
  The Public 4
  Health Professionals 5
  Industry 5
  Nature of the Challenge 6
Conceptual Framework for Medical Technology Assessment 7
Identification: Technologies Needing Assessment 8
Testing: Types of Information Needed and Mechanisms for Testing 10
  Health Effects 10
  Economic Effects 11
  Social Effects 12
  Mechanisms for Testing 12
Synthesis: Using Information as the Basis for Decisions 13
  Synthesis of Research Findings 14
  Synthesis of Health, Economic, and Social Effects 14
  Synthesis of Opinion 15
Dissemination of Information 15
  Government Activities 15
  Other Mechanisms 16
Major Conclusions of the Study 16
  Identification 16
    Emerging Technologies 16
    New Technologies 17
    Existing Technologies 17
    New Applications of Existing Technologies 17
  Testing 17
  Synthesis 17
  Dissemination 17
Policy Options 18
  Legislative Options 18
  Oversight Options 18
Organization of the Remainder of the Report 19

List of Figures

Figure No. *Page*

1. Process of Assessing Medical Technologies 7
2. Lifecycle of a Medical Technology 8

1.

Introduction and Summary

NEED FOR A STRATEGY

Several reasons for assessing medical technologies have been presented in previous OTA reports, *Assessing the Efficacy and Safety of Medical Technologies* and *The Implications of Cost-Effectiveness Analysis of Medical Technology*. The main reasons are to help ensure that medical technologies are safe, efficacious, and appropriately used. Whether current policies and practices for medical technology assessment achieve these and related objectives is the subject of this report. Having studied both the methods of medical technology assessment and the dissemination of information developed by technology assessment, OTA finds that a strategy is needed to implement the assessment process to make it more effective. OTA also finds that greater attention to assessment of social and ethical values is needed for policymaking.

A *medical technology*, as used in this report, is a drug, device, or medical or surgical procedure used in medical care. (The term may also apply to the organizational and supportive systems within which medical care is delivered, but those systems are not the focus of this report.) *Medical technology assessment* is, in a narrow sense, the evaluation or testing of a technology for safety and efficacy. In a broader sense, it is a process of policy research that examines the short- and long-term consequences of individual medical technologies and thereby becomes the source of information needed by policymakers in formulating regulations and legislation, by industry in developing products, by health professionals in treating and serving patients, and by consumers in making personal health decisions. Unfortunately, that process currently has deficiencies that cause or allow confusion to exist at all decision points.

Historically, medical technology assessment has developed incrementally as responses to specific demands. Taken singly, some of these responses have been coherent (e.g., the Food and Drug Administration's (FDA's) premarketing approval process which was developed to protect the public from unsafe and inefficacious new drugs). Taken in combination, however, these various responses do not constitute a coherent system for assessing all classes of medical technologies. The present approach is characterized by multiple participants from the public and private sectors, and by uncoordinated activities. Complicating matters further is the large number of medical technologies in use, with thousands of new technologies appearing every year. The result is an overload and confusion among decisionmakers and consumers.

OTA finds that a strategy is needed to guide the selection and implementation of components that would constitute a coordinated system of medical technology assessment. The basis of the strategy should be the values and available resources in a free-market economy, coupled with the social responsibility to make available safe, effective medical care. The vehicle of the strategy should be a systematic process of information development, dissemination, and use. The target should be to address the confusion deriving from the lack of information available to decisionmakers.

Minimally, the following components of an assessment system must be considered in developing a strategy:

1. the values of individuals and of society concerning medical technologies and their use;
2. the goals and appropriate role of medical technology assessment in society;
3. the types of assessment information needed for decisionmaking;
4. the methods and technologies for developing and acquiring the information; and
5. mechanisms for disseminating and applying the information, including programs that will use the information.

A strategy for assessing medical technologies must consider not only the methods of assessment, but also the needs, demands, and resistances of potential participants in the process of assessment. Specifically, the public itself as consumers; health care professionals as users; industry as innovators, producers, and reimbursers; and the Federal Government simultaneously as purchaser and guardian must be informed and active in setting mutually compatible goals for technology assessment. Each sector has health, social, and economic values underlying its decisionmaking. Clarifying those values and realistically accommodating them will require developing not just more, but also more reliable information about the safety, efficacy, cost effectiveness, and social and ethical implications of all classes of medical technologies. The inconsistencies and contradictions in available information are reflected in the inconsistent and competing pressures from the various sectors.

DIMENSIONS OF THE NEED AND THE PROBLEM

As an illustration of the mutual involvement of all sectors with a medical technology and as an illustration of the waste and potential threat resulting from premature adoption of an unassessed technology, consider the medical procedure of gastric freezing. In the mid-1950's, a clinician researcher at a university medical school, in conjunction with a private corporation, developed a device to treat peptic ulcer disease with gastric cooling. The procedure involved circulating alcohol at −15° C through a nasogastric tube to a balloon inserted into the stomach. He first tried the procedure on dogs, then on a dozen human patients, and reported in 1962 the following results: no serious side effects, reduced stomach acid output, immediate relief of ulcer pain, and radiographic evidence of ulcer healing. By the end of 1963, 1,000 devices had been sold, and 15,000 procedures had been performed nationwide, aided financially by third-party reimbursers. In 1964, other published reports concluded that acid suppression was limited or was unrelated to pain relief, symptomatic improvement was short-lived or due to placebo effects, and serious risks were present. By 1966, the technique was rarely used.

As an extreme example—that is, a technology that did not work—gastric freezing makes obvious the useless expenditure of money, time, and human emotion. The questions about most technologies, however, are more subtle. Most medical technologies have a therapeutic or diagnostic value for specific problems under appropriate circumstances. The difficulty is determining for whom and under what circumstances use of a technology is valid or worth the tradeoff of risks and benefits. Mammography and radical mastectomy, for example, have a place in the detection and treatment of breast cancer, but understanding exactly what that place is may take years and a certain amount of trial and error.

Government

The Federal Government's interest in developing clear policies and an effective strategy for assessing medical technology derives from its traditional role as guardian of the public's safety and of social equity and from its concerns about economic issues. As protector of the public, the Government seeks to ensure that health care is not only safe but also efficacious. As the single largest buyer of health services, the Government seeks to ensure that all citizens, especially the poor, have health care available to them; but the Government is also concerned about rising health care costs in general and specifically about those it pays for directly through programs of service or reimbursement (Medicare, for example) and through biomedical and other health research. Any policies the Government sets will affect not only the Government itself, but the public and private industry, and such policies must especially be justifiable when the public and private industry make self-interested demands.

The Public

The gastric freezing incident, though occurring 15 years ago, is still representative of current issues. As more recent technologies receive widespread attention (e.g., mammography, laetrile, and electronic fetal monitoring), the public becomes more vocal and involved. The public is often neither fully informed about the safety and efficacy of individual technologies nor educated about the issues of cost and social values that must be considered in the adoption of a technology. The public mixes facts with beliefs, hopes, and fears and translates those into confused, contradictory, and often impossible demands.

For example, the public hears of a drug, perhaps one used in another country, and wants it immediately available to patients in the United States, especially when available therapies are ineffective. The desperation individuals feel tends to outweigh the fear of any risks that might be involved, and they demand the right to take personal responsibility for use of the drug. Perhaps assuming that if a therapy is used in a European country, it has already met rigorous assessment standards, the public perceives itself as being denied a cure for no valid reason. The recent laetrile issue is perhaps the most emotionally dramatic example.

Simultaneously to demanding speedy availability and personal responsibility, however, the public demands protection against all forms of unsafe medical practice and is prepared to sue for mistakes. Perhaps because of the rigor which FDA applies to approval of new drugs and because of Government safety standards applied to so many nonmedical products, the public assumes that it is likewise protected in undergoing any medical or surgical procedure recommended.

The confused demands of the public can be viewed either as irrational or as a frustrated reflection of the deficiencies that do exist in the Nation's approach to assuring the availability of safe, effective, and cost-effective medical technologies. Numerous needs and values are implied in the demands of the public and must be taken into account when planning a strategy of policies and procedures for medical technology assessment.

Health Professionals

Health professionals often find themselves in circumstances that require decisions based on inadequate information. They take action or advise patients who must make decisions about use of drugs, devices, or medical and surgical procedures. Although at the time of decision the patient may be willing to assume responsibility for the decision, later, if harm occurs, the patient tends to hold the physician responsible. The flaws in the information flow to physicians and other health professionals are numerous: there is not enough information available about the safety, efficacy, costs, and social values of medical technologies; much existing information is of dubious quality and is therefore unreliable; the practical significance of data is usually not interpreted for clinicians; and easy access to the appropriate information is rare.

Furthermore, medical education typically does not train physicians and other health care professionals to make decisions based on a consideration of values. They are trained to seek the most reliable technique to produce a desired physiological response. As an illustration, in the issue of saving the lives of extremely premature babies in incubators, physicians, by training, would tend to be concerned mainly with choosing the technology that would support life. Physicians would less likely know or be concerned about the implications of the survival of the deformed or retarded infant—implications for the infant itself, for the family, and for society. Thus, developing and supplying the right kind of assessment information to health care professionals is essential to a strategy for medical technology assessment.

Industry

From the point of view of the private sector, of producers of technologies and of third-party payers, the assessment of medical technologies is both advantageous and disadvantageous. Government's involvement in the assessment process raises primarily financial issues for the private sector.

Industry, which invests money in research and development (R&D), is willing to do so if there is a potential market for the device or drug; however, excessive regulation or the wrong kind of regulation by the Government could discourage innovation if companies fear that assessment will ultimately preclude marketing their product or making a profit from it.

Private third-party payers, on the other hand, might welcome shifting the entire burden of assessment to the Federal Government. They must make decisions about reimbursement—whether to reimburse for specific procedures and if so how much—but they have little incentive to conduct their own assessments of procedures because of the expense. Assessment information tends to be widely available and not proprietary; the insurance companies cannot profit individually from conducting assessments. The failure of industry members to adequately conduct assessment activities on their own puts a heavy responsibility in the Government domain.

Nature of the Challenge

The market for medical technologies is moderated by individual consumer tastes and financial constraints. To perhaps a greater degree, it is influenced by policies that determine what kinds of research will be supported, what regulations restrict market entry, and which technologies will be reimbursable by Government or private programs.

No policy decision has isolated effects in just one sector; repercussions occur throughout the entire social and economic fabric of the Nation. A regulatory decision to require extensive, expensive assessment of a medical device in a developmental phase, but not to offer industry assistance in the assessment, for example, could lead to a decision by industry never to begin the innovation phase. An idea might never be realized which eventually could have best served the public. In fact, current policies and procedures for assessment are not adequate to fully serve the public interest. No consistent policy or system exists for assessing all classes of medical technologies, nor even for various technologies within a class.

The principles of competition and of supply and demand which ordinarily control prices and consumer choices in the market do not operate efficiently in the provision of medical care, especially because of reimbursement policies. Typically, for example, after a dramatic new procedure becomes routine, requiring less time and skill and incurring fewer risks, fees increase rather than decrease. Hospitals invest in services and new equipment which, like the hospitals themselves, are often underutilized. Third-party coverage of medical care, both Government and private, is a major cause of this inflated purchasing and cost. For this reason, the 1972 amendments to the Social Security Act limited the amounts Medicare could pay institutions and physicians.

Reimbursement decisions also influence the innovation and adoption of medical and surgical procedures. Although new procedures tend to be adopted and reimbursed without adequate assessment, in the case of truly innovative procedures, third-party payers sometimes refuse reimbursement. While encouraging new applications, slight modifications, and excessive use of existing technologies, the present reimbursement system may discourage radical innovations.

The challenge in developing a strategy for assessment is to develop a system that will serve the public interest by encouraging the development and appropriate use of needed and safe medical technologies without unnecessarily discouraging innovation and production. The practical questions are: What information is needed to make decisions about medical technologies in the best interest of the public and how can that information best be generated and disseminated? Can clear knowledge be developed that will enable policymakers and decisionmakers to act in the best interest of the social and economic elements of the Nation?

But there are also philosophical considerations. What the role of Government is and how strong that role should be is the subject of a perennial debate. Should the role be regulatory or oversight? To what degree? Should industry be left to its own incentives or pressured by the Government with directed incentives? How, in other words, can the Federal Government move the country toward a more efficient and equitable system of ensuring that useful and timely information is available to those who need it, without adversely affecting the innovation process and health care services?

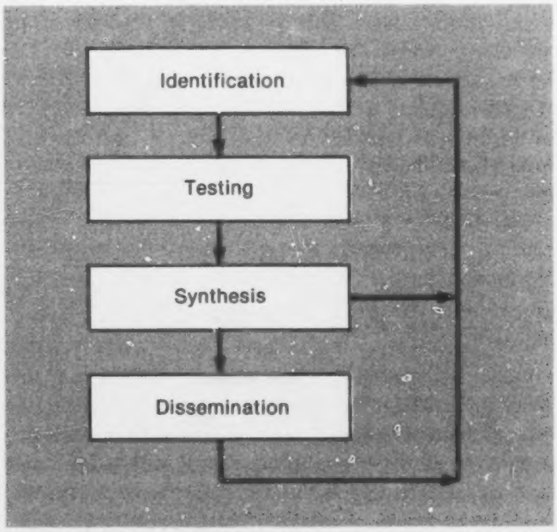
The next section of this chapter presents the major components, drawbacks, and considerations in the existing process of medical technology assessment.

CONCEPTUAL FRAMEWORK FOR MEDICAL TECHNOLOGY ASSESSMENT

Medical technology assessment involves numerous components and subcomponents at various stages of the process. Though these do not exist as a coherent system, discussion of them is facilitated by describing a systematic framework. The multiple components of the medical technology assessment process can be conceptualized as an information flow associated with the following four stages of assessment (see fig. 1):

- *Identification.*—Monitoring technologies, determining which need to be studied, and deciding which to study.
- *Testing.*—Conducting the appropriate analyses or trials.
- *Synthesis.*—Collecting and interpreting existing information and the results of the testing stage, and, usually, making recommendations or judgments about appropriate use.
- *Dissemination.*—Providing the synthesis of information, or any other relevant information, to the appropriate parties who use medical technologies or make decisions about their use.

Figure 1.—Process of Assessing Medical Technologies



SOURCE: Office of Technology Assessment.

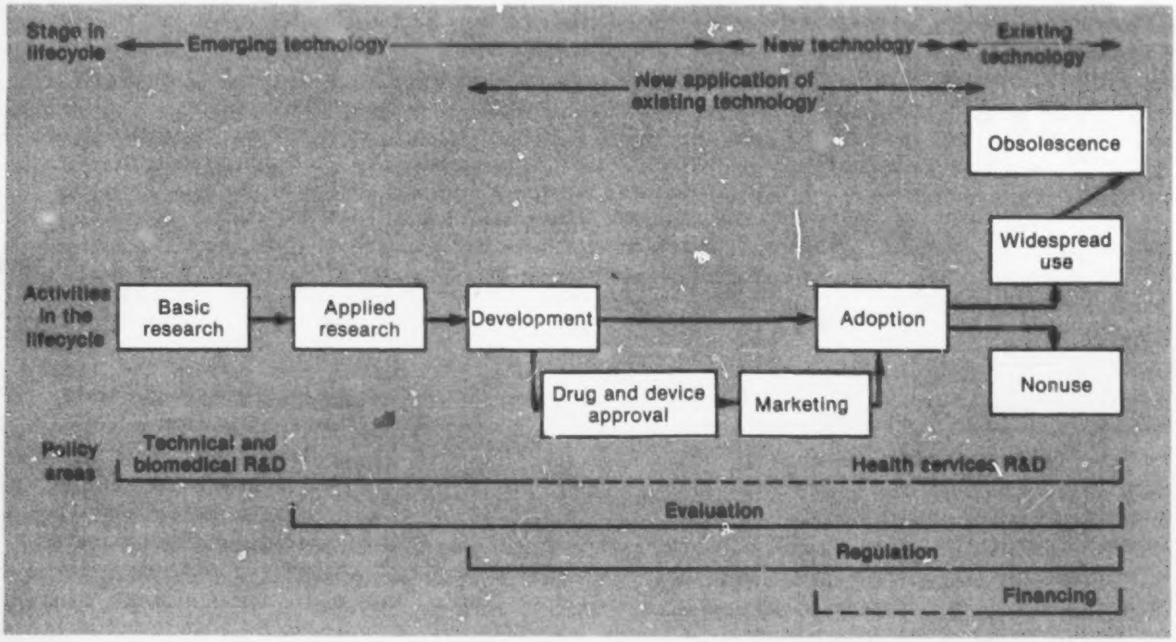
This four-stage process is applicable to the three classes of medical technologies mentioned earlier—namely, drugs, devices, and medical and surgical procedures. It is also applicable to any technology in any of four typical stages of development, loosely defined as follows:

- *Emerging technology.*—A technology in the phase prior to adoption.
- *New technology.*—A technology in the phase of adoption.
- *Existing technology.*—A technology in general use.
- *New application of an existing technology.*—A new application of a technology in general use.

Visualizing the lifecycle of a hypothetical technology (see fig. 2) makes obvious some of the decision points at which assessment information is essential. If an emerging technology is a drug or device, industry must decide whether to commit resources to develop it; must later decide whether to market it; and must ultimately decide whether to maintain, alter, or discard it. If a new drug or a certain class of device is to be marketed, FDA must decide whether to grant market approval based on safety and efficacy criteria. If the new technology is to be used in medical practice, someone must decide whether to pay for it. In some cases, the Health Care Financing Administration (HCFA) must decide whether to include a new technology or a new use of an existing technology as a reimbursable expense for Medicare beneficiaries. Private insurers, such as Blue Cross/Blue Shield and health maintenance organizations, must make similar decisions. Hospitals must decide whether to purchase, and practitioners and their patients must decide whether to use, the technology. Finally, all users and payers at times need to review the usefulness of existing technologies. And, in some cases, existing technologies find new uses or are modified, and the process begins all over.

In contrast to drugs and devices, medical and surgical procedures and their variations are ordinarily developed by clinicians and researchers and

Figure 2.—Lifecycle of a Medical Technology


Stage in lifecycle
Emerging technology
New technology
Existing technology
New application of existing technology
Obsolescence
Widespread use
Activities in the lifecycle
Basic research
Applied research
Development
Adoption
Drug and device approval
Marketing
Nonuse
Policy areas
Technical and biomedical R&D
Health services R&D
Evaluation
Regulation
Financing

SOURCE: Office of Technology Assessment.

therefore seldom require investment decisionmaking by industry. Furthermore, under the present system, medical and surgical procedures are not regulated for safety and efficacy by FDA and thus tend to escape the regulatory decisions. Nevertheless, decisions about reimbursement of such procedures must be made.

Many medical technologies in use have not been adequately evaluated. If all medical technologies were adequately assessed as emerging or new technologies, there would be less need for assessing existing technologies.

In addition to considering the stages of the assessment process and the classes and developmental stages of technologies, an assessment system requires the measuring of specified effects. Depending on the technology, the effects to be considered are health (safety, efficacy, and effectiveness), economic, or social. Once the categories of effects to measure have been determined, testing and analysis may begin. Throughout the assessment process, all information and decisions must be balanced against the moral and ethical values of society.

IDENTIFICATION: TECHNOLOGIES NEEDING ASSESSMENT

A decision to conduct a technology assessment must be preceded by the identification of technologies that should be assessed and the setting of priorities among candidate technologies. Identification procedures may vary with the type of technology, but basically can be classified as one of three types: 1) routine mechanisms, 2) priority-setting mechanisms, and 3) mechanisms of opportunity. Routine mechanisms systematically identify a class of technologies, usually in relation to a specific event—e.g., FDA requires that all drugs and devices be registered before thay can be marketed or tested in humans. Priority-setting mechanisms are used, as needed, to apply implicit or

BEST AVAILABLE COPY

explicit criteria to determine which technologies should be assessed—e.g., HCFA and the National Institutes of Health (NIH) set research agendas. Mechanisms of opportunity are not formalized but are valuable in identifying technologies as they surface or become important—e.g., patient outcome data may bring the need for analysis to the attention of researchers or the public.

Identifying medical technologies for priority-setting and assessment is an important responsibility primarily of several agencies within the Department of Health and Human Services (DHHS): FDA, the National Center for Health Services Research (NCHSR), NIH, and HCFA. The National Center for Health Care Technology (NCHCT), while it was funded, also identified technologies for assessment.

FDA identifies new drugs and medical devices through its premarket approval authority. To test promising new drugs in humans, drug sponsors (e.g., manufacturers) must notify and receive permission from FDA through a "notice of claimed investigational exemption for a new drug" (IND). If the drug successfully passes this premarket testing, the sponsor may file for a "new drug application" (NDA), which is a request for FDA's permission to market the drug. Since 1962, when this regulatory mechanism was instituted, FDA has reviewed over 13,500 applications for INDs and has approved about 1,000 NDAs. Since 1976, FDA also has an expanded responsibility for regulating medical devices. In the first 4 years of implementing the 1976 Medical Device Amendments, about 98 percent of the listed devices in the 10,540 premarket notifications received were claimed to be "substantially equivalent" to preexisting devices. In 1981, FDA estimated that 2,300 premarket notifications would be reviewed. New applications of existing drugs and devices must also meet premarket approval requirements, but the initiative for these new applications remains with the manufacturer, not with FDA. FDA does support some monitoring activities of existing drugs and requires manufacturers to report adverse reactions, but these postmarketing surveillance activities are focused on the safety aspects of these drugs, not on refinements in use or new uses.[1] Nevertheless, postmarketing surveillance has the potential of becoming an effective method of identifying existing technologies in need of further assessment.

NIH and *NCHSR* are research agencies that identify emerging and sometimes new technologies in need of assessment through their priority-setting processes for research grants and contracts. Projects are selected on the basis of technical merit and whether they are addressing important issues. The processes generate information useful to policy decisions, but do not necessarily address the immediate priorities of operating agencies such as HCFA.

HCFA reimburses for Medicare and therefore has obvious incentives for identifying technologies in need of assessment; nevertheless, it has no mechanisms for the identification of existing technologies in widespread use. For new technologies, the identification is by opportunity. When the question of coverage arises for new technologies, HCFA must determine whether it has adequate information to make a decision and must set priorities for technologies that must be assessed to provide more information. Also, through its Office of Research and Demonstrations, HCFA sets priorities for assessing technologies that are important to its operations.

NCHCT was established in 1978 to undertake and support assessments of medical technology, but did not receive funding in 1982. NCHCT coordinated interagency issues, but also set its priorities internally and had its own responsibilities for identifying technologies. Specifically, NCHCT compiled an annual "emerging technology list" as an early alert system for assessment, but the 1981 reauthorization of NCHCT withdrew its authority to compile the list. (Industry argued that the list threatens innovation by casting doubt on the eventual marketability of a technology.) NCHCT also initiated a plan to develop a joint public-private model for collecting clinical data on emerging technologies. Finally, the NCHCT Director chaired the Technology Coordinating Committee of DHHS, which was the department's primary mechanism for coordination of issues associated with medical technologies.

Overall, the current identification stage of the current system of technology assessment has serious shortcomings. The degree to which current processes identify technologies varies. Emerging and new drugs and devices are adequately identified for assessment prior to their being marketed. However, emerging and new medical and surgical procedures are not adequately identified, because no one in either the private or the Government sector has a clear responsibility for the task. New mechanisms are especially needed to identify for the purpose of assessment existing technologies of all classes, new applications of existing technologies of all classes, and medical and surgical procedures in all four phases of development.

---

[1] This topic is explored in greater depth in OTA's report entitled *Postmarketing Surveillance of Prescription Drugs.*

TESTING: TYPES OF INFORMATION NEEDED AND MECHANISMS FOR TESTING

As a basis for decisions, a strategy to assess medical technologies must take into account what is known, what is not known, what is needed, what can be obtained, and at what cost. Information will never be perfect, and money and time will always be limited; thus, evaluation methods must be used judiciously and their results must be interpreted cautiously, in conjunction with numerous other measurements, especially with consideration for society's moral and ethical values. Three categories of information about a medical technology are needed for policy decisions: 1) health effects, 2) economic effects, and 3) social effects. The methods and procedures for determining these effects have strengths and weaknesses closely paralleling those of the identification phase.

Health Effects

Health effects are determined during the testing stage of assessment. The basic questions asked are: Does the technology work? and How well does it work? The former question seeks information about efficacy, effectiveness, and performance standards, and the latter about safety (and risk). The information provided by analyses of health effects helps decisionmakers determine whether a drug or device should be allowed on the market or whether further investment in R&D is warranted.

Patient outcome is the desired endpoint measured in efficacy and effectiveness analyses; efficacy is tested under ideal clinical conditions, whereas effectiveness is tested under average, or typical, conditions. Tests for effectiveness demonstrate whether efficacy information can be generalized to the population at large. For new drugs and certain devices, if the technology is in the emerging phase, its efficacy must be established in preclinical, biochemical, or animal tests before it can be tested among humans. The method that gives the most valid and most reliable information about efficacy is the randomized clinical trial (RCT). The strength of the RCT lies in its randomization process, producing two or more groups that are identical except for chance occurrence, which can be estimated statistically. The drawbacks of RCTs are that they can only be used in certain settings, they are sometimes not ethical to conduct, and they do not always provide complete information about safety.

Thus, despite the highly valid information they can produce, RCTs are not always the method of choice. Other methods can be used as substitutes for RCTs or to supplement them. Observational methods are designed to analyze data from nonrandomized study designs. Several techniques are used to minimize selection bias. Observational methods can be useful in ruling out competing explanations for an observed effect and for testing hypotheses in large, diverse populations once a technology is widely diffused. Prospective cohort studies, for example, can be used to detect rare adverse reactions to drugs that were unsuspected prior to marketing. Case-control studies are an inexpensive means of indicating whether the use of a technology results in a small level of risk.

Another, more common type of study is the case study, typified by a physician reporting his or her experience with particular technologies and patients. Case studies are useful in an overall assessment strategy in that they can facilitate the identification of technologies in need of assessment. Case studies are important identification mechanisms of opportunity, as defined earlier. However, the validity of case studies is extremely low because of, among other things, observer bias and the placebo effect. Nevertheless, clinicians are very often swayed by these case reports, which fill the medical literature and which often describe the successful application of a technology.

Safety is measured in terms of a risk-to-benefit ratio; it is therefore a relative concept, and its estimation may be a byproduct of testing for efficacy and effectiveness. A low risk may be unacceptable if there is no benefit, but a high risk may be acceptable if the benefits are also high. RCTs tend to give risk information only on a small segment of the population. To generalize to other segments, supplemental information is needed from surveys and methods which can make use of registries, and clinical data banks.

For certain technologies, especially devices, establishing the technology's performance integrity is a prerequisite for efficacy assessment. Performance standards usually pertain to the chemical, physical, and electric properties of devices. Similar standards are often used in evaluating technologies which have an intermediate rather than a direct effect on the patient's health outcome, e.g., diagnostic and often prevention technologies. In such cases, the technology is evaluated in terms of its ability to cause one effect that in turn will cause the desired result. For example, an automatic blood pressure monitoring device must accurately measure and record blood pressure if it is to be used for diagnostic purposes. Coronary artery bypass surgery is a preventive procedure for heart attack in that it increases blood flow to the heart, the expectation being that pain and the likelihood of a heart attack will be reduced.

No precise formula exists for choosing the best or most appropriate evaluation method. The stage of development of the technology itself—e.g., emerging, new, or existing—will partially determine the appropriateness of a method. The purpose of the technology—e.g., diagnostic, therapeutic, or preventive—will limit the range of appropriate methods. However, other factors such as existing knowledge about the risks and benefits and available resources may influence or override otherwise "ideal" choices. The important criterion in selecting analytic methods is not which is theoretically more sophisticated, but which is practically the most appropriate.

Economic Effects

What does it cost the Nation and the individual to develop and use a medical technology? What does it cost *not* to develop or use a specific technology? The answers to these questions supply decisionmakers in Government and industry with information they need for allocating financial resources. All who pay for care—Government, insurance companies, individuals—need to know whether the use of the medical technology is worth the cost.

Analytical methods to determine economic effects comprise a spectrum ranging from sophisticated computer-based data analyses to best-guess estimates of costs and benefits. The broad terms cost-benefit analysis (CBA) and cost-effectiveness analysis (CEA) refer to two techniques for comparing the positive and negative consequences of alternative ways to allocate resources. The principal distinction between the two is that CBA values all costs and benefits in monetary terms whereas CEA produces a measure of the cost involved in terms of some desirable health-related effects (e.g., years of life gained).

Measurements of economic effects should consider both direct and indirect costs. Direct costs are those associated with direct medical care usage: the cost of the physician, the hospital, the medical supplies. Indirect costs are associated with the value of time lost in receiving medical care and in being sick. When indirect costs are considered in economic analyses—and often they are not—they are frequently measured in terms of lost or gained wages.

Economic analysis is complex and must consider more than charges for services. For example, cost analyses should develop information on opportunity cost, marginal valuation, joint production considerations, R&D costs, overhead, costs v. prices, and discounting. Just as no one method is invariably appropriate in the evaluation of health effects, no one method of economic analysis is appropriate. The user of the information will partially determine the kinds of analyses done. For the patient, the actual cost of services is the important information. For policymakers, more complex information is required. In the sequence of the assessment process, information about the economic effects may be useless if reliable and appropriate information about the health effects is not available.

Social Effects

Urgent ethical and social questions are being raised in areas of biomedicine such as experimentation with human subjects, genetic engineering, human reproduction, and the possibly inappropriate prolongation of life. Who is affected by a medical technology? Who is not affected? What values of individuals and society are involved in use of the technology? What ethical principles are involved in testing the technology?

To varying degrees, medical technologies may affect the personal and work lives of patients and their families; influence the structure of medical, legal, and economic systems; and challenge society's most fundamental beliefs. Considerations of the social and ethical implications of medical technologies, therefore, must take an important place in the development of policies. Social implications are the direct or indirect effects of medical technology on the concepts, relationships, and institutions society considers important. Ethical questions in relation to medical technologies—especially those concerning principles of distributive justice, respect for individuals, and benevolence—may also have profound social implications.

Unlike health and economic effects, social and ethical issues do not lend themselves to quantitative measurement and analysis. However, the systematic identification and evaluation of the social impacts resulting from the use of medical technologies can be crucial. A related task is to identify the values that underlie policy alternatives, including moral and ethical values. Systematically assessing values does not necessarily elucidate a single, clear, conclusive answer about which policy to adopt; but, rather, it clarifies the array of choices, the reasons for disagreements, and the compromises required.

A second aspect of assessing values is to make a reasonable inquiry into the values that permeate and underlie the assessment itself. Value judgments enter into every aspect of technology assessment; they determine which technologies will be assessed and at what phase of their development, the scope of assessments, the kinds of data that will be collected and analyzed, the methods of the assessment, and how the assessment findings will be used in decisionmaking. It is important to clarify, therefore, why an assessment of a particular technology was initiated and how it fits into larger cultural and political contexts, what affects the performance of assessment (e.g., the choice of assessors and the analytic goals and methods), and what values affect the application of the results.

Mechanisms for Testing

The major problem with the testing phase of the current assessment system is the lack of a systematic approach for testing identified technologies in all phases of development for all types of required information.

FDA, in its regulatory role, is probably the most significant agency in stimulating technology testing. Most FDA regulation requires industry to test, according to approved protocols, new drugs and many medical devices for safety and efficacy. For drugs, Phase I studies determine levels of tolerance (toxicity), followed by early dose ranging studies for safety and sometimes efficacy. If safe, the drug can be tested in Phase II studies to demonstrate efficacy and relative safety under controlled conditions. Phase III studies are expanded controlled and uncontrolled clinical trials. If these trials are successful, the compnay may file an NDA. FDA then reviews the data and may approve the drug for marketing. Since 1962, FDA has approved about 1,000 NDAs. For devices, FDA requires that 90 days notice be given about any new device industry intends to market. If a device does not meet safety and performance standards for its assigned classification, or if adequate information is not available for such a determination, FDA may require testing of the device. For drugs and devices, FDA's assessment activities are generally limited to safety and efficacy and do not involve cost, cost effectiveness, or social effects.

Unlike drugs and devices, medical and surgical procedures are not regulated, and their testing, if done, is through research whose funding comes primarily from NIH and from private foundations. The costs of the later developmental phase of procedures tend to be paid by patients (or by the Government), usually through standard medical insurance policies, even when the procedure has been clearly designated as experimental. Medical and surgical procedures usually begin as user-generated innovations; for example, a surgeon may modify an existing technique during surgery. Increasingly, innovations arise in academic centers, from researchers who know how to present their innovations in a technically acceptable manner at professional meetings and in journals. These researchers' presentations tend to legitimize innovations without their receiving a routine, formal examination for safety and efficacy.

Whereas FDA regulations affect efficacy and safety, four other regulatory programs are concerned with cost issues: section 1122 review, State certificate-of-need laws, the National Health Planning and Resources Development Act of 1974, and Professional Standards Review Organizations (PSROs). Although HCFA, which makes reimbursement policy, has its own research arm, the Office of Research and Demonstrations, it has seldom conducted technology assessments. NCHCT, an agency legislatively mandated to support comprehensive assessments of health care technologies for all effects (including health, economic, and social), was not funded for 1982.

Social assessment activities have been conducted by several Government mechanisms. OTA was established in 1972 as an analytic support agency to conduct policy research on science and technology issues for congressional committees. OTA's health-related reports have focused primarily on methods available for assessing technologies and issues prompted by their use. The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research was established in 1974 to develop ethical guidelines for conducting research in human subjects. The National Commission produced numerous reports with recommendations, many of which were adopted by DHHS,* particularly those governing the protection of human subjects. The Ethics Advisory Board, which was established in 1978 at the National Commission's recommendation but was not funded in 1980, was mandated to review ethically problematic research protocols and research involving human projects. The board fielded queries from other DHHS agencies such as NIH and the Centers for Disease Control. The President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research succeeded the National Commission in 1978. Members of the President's Commission are appointed representatives from DHHS, the Department of Defense, the Veterans Administration, the Central Intelligence Agency, the National Science Foundation, and the White House Office of Science and Technology Policy. The President's Commission conducts studies in medical practice and biomedical research and examines five subjects for legal and ethical implications: informed consent, privacy, uniform definition of death, genetic issues and unborn humans, and availability of health services. NCHCT's responsibilities, as mentioned above, included assessment of the ethical, legal, and social implications of medical technologies.

---

*Then the Department of Health, Education, and Welfare.

SYNTHESIS: USING INFORMATION AS THE BASIS FOR DECISIONS

Synthesis of the information generated during the testing stage of the assessment process is the necessary step to providing a convincing and responsible basis for decisions made during all phases of a technology's lifecycle.

The synthesis activities that pertain to medical technology assessment fall into two broad areas: 1) synthesis of the results of individual research studies; and 2) synthesis of a body of research findings with various concerns such as risk, social, ethical, or cost factors. The first type of synthesis addresses questions of safety, efficacy, or effectiveness of a given technology; the latter is more policy oriented, often seeking to set guidelines or standards for medical practice or reimbursement policy. The value of the latter depends, in large part, on the adequacy of the former.

Synthesis of Research Findings

The traditional approach to synthesizing research information is the literature review, an article summarizing the data of those studies a reviewer believes to be the most relevant to the topic under review. Literature reviews are useful and heavily relied on, but because of their scope and the delays in the journal publication system, such reviews are rarely timely, especially in reporting an ineffective or unsafe technology. Furthermore, the reviews are subjective and often have no commentary on methodological problems in individual studies.

More systematic procedures for integrating and interpreting sets of research evidence do exist and can be employed. The most simple technique is a simple classification technique, sometimes called the "voting method." This technique involves selecting a sample of evaluative studies, coding some aspect of the design, classifying outcomes as favorable, neutral, or unfavorable and constructing tables of research findings. The method identifies methodological strengths and weaknesses among studies and can help determine patient populations and under what conditions they are most likely to benefit from a technology.

Meta-analysis is a technique that assesses the magnitude of treatment impact by quantitative comparison of actual study results. This method is particularly useful in assessing treatments for which a large number of studies are available and findings across studies seem to have great variability. However, it may have drawbacks with respect to sample selection.

Currently, no single technique is fully adequate for synthesizing research; however, the application of formal quantitative procedures is beginning to give a better understanding of methodological problems in research itself. Formal procedures can segregate differential outcomes according to treatment characteristics and methodological approaches. Contradictions can then be identified, analyzed, or further researched. In the performance of formal quantitative analyses, an important suggestion is that the significance of the results should be interpreted and reported in language that is useful to decisionmakers.

Synthesis of Health, Economic, and Social Effects

How, then, does one bring together and synthesize all information available about all three categories of the effects of medical technologies—health, economic, and social? Once specific information has been synthesized through various methods in each of these realms, how can a decisionmaker balance the values and interpret them into programmatic actions?

OTA's report on CEA concluded that performing an analysis of costs and benefits can be very helpful to decisionmakers, because the process of analysis gives structure to a problem, allows an open consideration of all relevant effects of a decision, and forces the explicit treatment of key assumptions. Formal techniques such as CEA can be used to aid in the synthesis of information concerning the health and economic effects of a technology. OTA found, however, that although CEA can be useful as a decision-assisting tool, it exhibits too many methodological and other shortcomings for the numerical results to be used as the basis of policy or program decisions. For example, although CEA can be used to synthesize information concerning health and economic effects, it cannot in itself adequately address social and ethical issues. These have to be addressed more fully by other means.

The most appropriate approach to any assessment is to perform it in an open forum so that assumptions and underlying values can be challenged; to identify, measure, and, to the extent possible, value all relevant benefits/effects and costs; and to present the results of the analysis as an "array" of benefits/effects and costs rather than forcing the results into a single aggregate measure. By arraying effects in a systematic fashion, one can place the appropriate relative emphasis on given effects whether they are quantifiable or not. This technique is designed to make more explicit the health, economic, and social consequences of any decision.

Synthesis of Opinion

Synthesis of information may occasionally present a clear-cut indication of the next stage of assessment or phase of technology development. More likely, uncertainty will still predominate for decisionmakers. The uncertainty may reflect the presence of random events or may reflect a basic lack of knowledge. The former can be analyzed by various statistical techniques: decision analysis, confidence limits, computer simulation, sensitivity analysis. However, these techniques cannot actually resolve policy controversies or substitute for informed judgment.

Policy judgments may require a synthesis of opinion which can be solicited from groups and expert input. The most common format of soliciting group opinions is the unstructured conference which may involve presentations, discussions, and debates. Another informal technique is the advisory panel approach used by many Government agencies. The four best known formal techniques used in medical contexts for resolving conflicts and uncertainty are: 1) the Delphi technique, 2) the nominal group process technique, 3) the consensus development conference (NIH), and 4) a computerized knowledge base which maintains expert opinion on the state of the art of a specific topic (e.g., the Hepatitis Knowledge Base of the National Library of Medicine, NLM). Although these formal techniques produce more reliable opinion information than an unstructured conference does, evidence of effectiveness is contradictory for the Delphi and nominal group processes and sparse for the NIH and NLM processes.

DISSEMINATION OF INFORMATION

What potentially are the direct effects of dissemination of assessment information? Who should have top priority in receiving information? How should the information be disseminated? The dissemination of assessment information directly affects the development and diffusion processes of medical technologies. The consideration of whether to disseminate information is therefore weighty. If a decision is made to disseminate information because the technology is deemed either worthy or unworthy of its next phase of development, the information must reach, at a minimum, the decisionmakers involved with the technology in any aspect of its use. That audience may range from directors of R&D in private industry, to health professionals, to the general public. Reaching the audience in a timely manner requires a systematic approach to information dissemination, especially in view of the pace and quantity of information development and the lack of mechanisms for the systematic synthesis of information. In a sense, the information available is at once too much and too little.

The dissemination phase of medical technology should comprise the mechanisms and coordination of communication activities. Unfortunately, current procedures are highly flawed; there exists no system for disseminating information, only a variety of traditional mechanisms. Little is known about the adoptive process or how information is used once it is received, but it is clear that medical practice varies greatly from provider to provider and that even when good information is available, many technologies are used inappropriately.

Government Activities

The Federal Government produces, collects, and disseminates assessment information. NCHSR, for example, disseminates the results of health services research to relevant Government agencies, the research community, and other interested parties through publications, press releases, conferences, and workshops. In 1978, the legislation authorizing NCHSR was modified to require that at least $1 million or 5 percent of its budget, whichever is less, be used for dissemination activities. In response, NCHSR established a User Liaison Program to provide substantive assistance to non-Federal health care leaders concerned with critical policy issues and operational problems in the organization, administration, regulation, and delivery of health care services at State and local levels.

Monitoring NIH's dissemination activities is the responsibility of the Office for Medical Applications of Research (OMAR), established in 1978 in the NIH Office of the Director, and assisted by the OMAR Advisory Committee. One important mechanism for dissemination is the consensus development conference. The synthesis of opinion that is achieved at a consensus conference is presented in consensus statements and supporting materials which are distributed to practicing physicians, other health professionals, the biomedical research community, and the public—through a mailing list of over 21,000 names. Also, members of the press are invited to the conferences and are encouraged to publish the results. Leading medical journals and medical societies have published the consensus materials.

In conducting medical technology assessments, information from several subject areas is often required. A common need in most assessments, however, is for information from the field of biomedicine. NLM is the major Federal library resource for biomedical literature. It is the predominant creator and disseminator of biomedical bibliographic information. NLM's coverage of the health services literature is less comprehensive than its coverage of the biomedical literature, in part because relevant health services information appears in so many diverse documents.[2] Another source of information for medical technology assessments is the National Technical Information Service (NTIS). NTIS is the central repository for scientific and technical information generated by federally funded R&D projects, including those in DHHS.

Other Mechanisms

Apart from formal Federal agency activities, mechanisms for dissemination include the public media, the mail, advertising, personal contacts, the educational process, libraries, and other types of information centers. The appropriateness of any of these mechanisms depends on whether the information is to be used in assessing or marketing a medical technology. Print media, radio, and television are primary channels to the public. In addition to news about medical technologies and issues, they increasingly tend to have health columns and special in-depth features about health technologies. For more targeted audiences, mailings are used for solicited and unsolicited information dissemination, for example, newsletters from drug companies, advertisements from product distributors, and Federal literature. Advertising of drugs occurs in all media for the public and for health professionals. A recently developed form of advertising, the video cassette, is supplied to medical facilities. Personal contacts are an especially credible source of information exchange among health professionals. These often occur formally and informally at professional meetings.

---

[2] This topic is explored at greater length in a separate OTA technical memorandum entitled *MEDLARS and Health Information Policy*, to be published in fall 1982.

MAJOR CONCLUSIONS OF THE STUDY

In this study of medical technology assessment, OTA has reviewed the evidence and concludes overall that there is no coherent system of assessing medical technologies. There is, however, an urgent need for such a system. The following are capsule statements of OTA's conclusions about the adequacy of the present system with respect to the four stages of technology assessment presented in figure 1: identification, testing, synthesis, and dissemination.

Identification

Emerging Technologies

OTA concludes that emerging drugs and devices are adequately and appropriately identified, but that emerging medical and surgical procedures could be better identified. Overall, however, the identification of emerging technologies for assessment is not a critical weakness of the present assessment system.

New Technologies

OTA concludes that new drugs and devices are adequately identified for the purposes of assessment, but that new medical and surgical procedures are not. The most pressing need is for some routine mechanism, e.g., the reimbursement system, to identify new procedures before they are widely adopted. The reimbursement system may be the prime candidate, because coverage and payment decisions are critical points in the diffusion of many technologies. The priority-setting systems of the institutes of NIH and of other Federal research agencies (e.g., NCHSR) are adequate and appropriate for their respective mandates, but there is not an adequate similar system to fulfill the needs of operating agencies (e.g., HCFA, planning agencies). Finally, sufficient mechanisms of opportunity for identifying new technologies could be developed. Medical specialty societies could be helpful in this area.

Existing Technologies

OTA concludes that the system for identifying existing technologies in need of assessment is inadequate. The most promising possibility for identifying such technologies may be FDA's postmarketing surveillance of marketed products. In the case of existing as well as new technologies, the priority-setting procedures of Federal research agencies may be adequate for those agencies' respective needs; however, these procedures are not adequate for the needs of operating agencies such as HCFA. And the operating agencies themselves do not adequately identify existing technologies for assessment. Medical specialty societies could be helpful in this area. Finally, NCHCT's activities of identifying nationally important priority technologies for assessment were valuable but are not currently funded. Thus, no organization is currently performing this important task.

New Applications of Existing Technologies

OTA concludes that new applications of existing technologies in need of assessment are not adequately identified. The most promising approach would seem to be the use of the reimbursement system to link the diagnosis with the use of technology. Medical specialty societies could be helpful in this area.

Testing

OTA concludes that, in general, drugs and devices are adequately tested for safety and efficacy prior to being marketed. Medical and surgical procedures, which often include the use of drugs and devices within the practice of medicine, are not well tested for either safety or effectiveness. No class of technologies is adequately evaluated for either cost effectiveness or social and ethical implications. Finally, there is no organization whose mission it is to ensure that medical and surgical procedures are assessed for safety and efficacy or to evaluate medical technologies for cost effectiveness and for social/ethical effects.

Synthesis

OTA concludes that the synthesis phase of the present system of technology assessment is unnecessarily weak, within both the private and public sectors. Research evidence regarding the safety, efficacy, and effectiveness from the use of medical technologies is seldom examined systematically and objectively. Federal agencies and private insurers and organizations set policies, guidelines, regulations, and/or make reimbursement coverage determinations, many of which profoundly affect the adoption and level of use of medical technologies. Yet, their decisions are usually based on informal, subjective, group-generated norms which tend to support the status quo. Formal, more objective techniques do exist, however, not only for evaluating research evidence but also for making decisions and setting policy. These techniques could be used more often to aid in better decisionmaking.

Dissemination

OTA concludes that better methods need to be found to communicate information about medical technologies to health practitioners, health researchers, and health policymakers.

OTA also concludes that Government-generated research reports, many of which may be important to technology assessment, are not as accessible as they could be. Finally, NLM's mission and capabilities should be examined to determine whether more Government reports and nonserial literature should be included in its data base, and whether NLM should index articles differently for researchers interested in technology assessments.

POLICY OPTIONS

The most important policy need is to bring forth a rational, systematic approach from the present multiplicity of agencies and activities to promote and coordinate medical technology assessment. Such integration could be accomplished in any of several ways. The options listed below and discussed at greater length in chapter 8 are divided into two broad categories: legislative and oversight. OTA finds that there are relatively few realistic legislative options necessary for Congress to consider, primarily because there is already substantial power invested in the Secretary of Health and Human Services to develop a coherent system of medical technology assessment. Thus, in most of the deficient areas noted within this report, congressional oversight may be sufficient.

Legislative Options

1. Sponsor or grant a charter to a private/public organization to undertake medical technology assessment activities.
2. Maintain the authority of, and appropriate funds for, NCHCT.
3. Change the statutes so that HCFA can selectively reimburse for experimental technologies in return for clinical data.
4. Increase funding to train researchers in methodological and statistical principles.
5. Increase efforts to train health professionals in methodological and statistical principles.

Oversight Options

6. Encourage the private sector to take the lead in assessing medical technologies.
7. Examine how Federal research institutes (e.g., NIH), agencies (e.g., NCHSR), and research programs of operating agencies within DHHS could identify technologies better when setting research agendas; and how the PSRO program and the reimbursement system could be used to more advantage for identifying technologies for assessment.
8. Continue to conduct oversight hearings concerning the duplication and fragmentation of health-related data collection activities.
9. Examine the ability of operating agencies within DHHS (e.g., HCFA) to generate sufficient information for their own decisions related to medical technologies, and examine the extent to which the Secretary of Health and Human Services utilizes the department's other research arms (e.g., NCHSR, NIH) to procure that information in a timely manner.
10. Examine the activities, plans, and potential for elements of DHHS (e.g., NIH) in utilizing various research methods to determine the appropriate use of medical technologies.
11. Explore how research evidence could be better evaluated by Federal health agencies when recommending, setting, or implementing health policy.
12. Examine the disposition of federally generated reports to determine how accessible and useful they have been both to private and public researchers and policymakers.
13. Examine whether NLM should include more Government research reports and other nonserial literature in its MEDLARS data bases.
14. Encourage use of the powers vested in the Secretary of Health and Human Services to develop a coherent system of medical technology assessment.

ORGANIZATION OF THE REMAINDER OF THE REPORT

Chapters 2, 3, and 4 discuss the types of information technology assessment seeks to generate, establish, and synthesize: namely, information on health, economic, and social/ethical effects. The methods and mechanisms used to synthesize that information are discussed in chapter 5. Chapter 6 includes a description of the drug and device industries, as well as a description of the innovation process for drugs, devices, and medical and surgical procedures. It also presents an analysis of the effects that reimbursement and Federal regulatory policies exert on the innovation process. A critique of current assessment policies and procedures in chapter 7 summarizes the strengths and weaknesses in each of the four stages of assessment and presents OTA's major conclusions. Chapter 8 presents the policy options.

Eight appendixes are included to serve as extensive technical data supporting and amplifying the issues and conclusions of the report. Appendix A and B are a compendium of statistical data sources for medical technology assessment and a compendium of bibliographic data bases for medical technology assessment, respectively. Appendix C is a paper on the methods used in the evaluation of medical technologies. Appendix D describes the innovation process for medical technologies, which five case studies in appendix E are intended to illustrate. Appendix F presents a proposed model for an Institute for Health Care Evaluation. The method of study and the other volumes of this assessment are described in appendix G, and acknowledgments appear in appendix H. Appendix I is a glossary of acronyms and terms.

Throughout this study, OTA paid special attention to the innovation process for medical technologies, since a successful strategy of assessment should not, at a minimum, unnecessarily interfere with beneficial innovation and, to the extent possible, should encourage useful innovation. OTA believes that none of the policy options presented in this report would unduly restrict the innovation of medical technologies.

Three other volumes are being published in conjunction with this report: 1) *Postmarketing Surveillance of Prescription Drugs,* 2) *MEDLARS and Health Information Policy,* and 3) *Medical Technology Under Proposals To Increase Competition in Health Care.* These volumes are briefly described in appendix G. In addition, chapter 1 of this report is available as a summary pamphlet.

2.

Information for Assessment

Knowledge is of two kinds. We know a subject ourselves, or we know where we can find information upon it.

—*Samuel Johnson*

Contents

	Page
Introduction	23
Information Needed for Assessment	23
Health Effects	23
Performance Standards	23
Economic Effects	24
Social Effects	24
Sources of Data	24
Data Banks	25
Vital Statistics	25
Insurance Claims Records	26
Surveys of Patterns of Medical Practice	26
Publications and Bibliographic Data Bases	27
Libraries and Other Information Resource Organizations	28
Conclusion	30

2.
Information for Assessment

INTRODUCTION

A primary consideration in developing a strategy for assessing medical technologies are the needs of assessors for various types of information. This chapter is an introduction to the information needed for assessing health effects, performance standards, economic effects, and social effects of medical technologies. (Subsequent chapters will describe and critique the methods for obtaining and synthesizing this information). The second section of this chapter also provides an overview of existing sources of statistical data for assessment, both public and private. The third section describes existing biomedical literature sources, including bibliographic data bases. And the final section discusses libraries, clearinghouses, and similar organizations which can provide information that is useful for technology assessment. To supplement the material in this chapter, a compendium of statistical data sources and a compendium of bibliographic data bases are included as appendix A and B.

The discussion of the sources of information for medical technology assessment in this chapter is not an exhaustive one. Because of the complexity and diversity of medical technologies, information from disciplines such as engineering, social behavior, and genetics is often needed to perform a technology assessment. The focus in this chapter is on the significant sources of information that is directly related to health.

INFORMATION NEEDED FOR ASSESSMENT

Health Effects

The information needed for a complete assessment of a technology's health effects falls into three broad categories: efficacy, effectiveness, and safety. *Efficacy* refers to the probability of (usually health) benefit to individuals in a defined population from a medical technology applied for a given medical problem under ideal conditions of use (266). *Effectiveness* is similarly defined, except that it refers to the probability of benefits under average conditions of use (266). *Safety* is a judgment of the acceptability of risk (i.e., the probability and severity of an adverse effect) associated with the use of a technology.

Information regarding the efficacy of a technology is needed to establish, within some defined estimate of probability, whether the use of a particular medical technology under "ideal" conditions can cause changes in patient outcome. Because information concerning the efficacy of medical technologies (under ideal conditions of use) often cannot be generalized to wide populations receiving medical care in diverse settings, information on the effectiveness of such technologies (under average conditions of use) is also needed.

Most medical technologies have an element of risk associated with their use. Thus, any assessment policy needs mechanisms to determine the probable risk, or, conversely, safety, of a technology under various conditions of use, and then to weigh the risk with the expected benefit. Weighing risks with benefits is necessary, because the concept of risk is a relative one—i.e., a low risk is unacceptable if there is no expected benefit, but a high risk may be quite tolerable if the expected benefit is very high.

Performance Standards

Strictly speaking, only an improvement in patient outcome can be considered evidence of a technology's efficacy or effectiveness. However, some technologies' ultimate effect on health may be so far removed from the technology's use that only some intermediate outcome can reasonably be assessed (270). For these technologies, information is needed concerning standards of performance.

Such information is especially needed in the case of medical devices, many of which perform some particular mechanical, electrical, or chemical function ultimately intended to be related to a change in health status.

In an earlier report, OTA classified relevant outcomes of diagnostic technologies to be: 1) technical capability, 2) diagnostic accuracy, 3) diagnostic impact, 4) therapeutic impact, and 5) patient outcome (266). Since diagnostic technologies are directly associated with intermediate outcomes and are often far removed from actual health outcomes, performance standards may be the most meaningful, as well as most easily obtained, measures of assessment.

Economic Effects

Good economic data are essential to a system of medical technology assessment. This is especially true today, when one of the driving forces behind technology assessment is concern for the high costs associated with the adoption and use of technology. For the purpose of assessing medical technologies, information is needed on the direct costs associated with medical care usage: the cost of the physician, the hospital, and medical supplies. Also important to consider are indirect costs: the costs associated with the value of time lost while seeking and receiving medical care and, especially, in being sick. Indirect costs are often overlooked in assessments; when included, they are often measured in terms of lost (or gained) wages.

Social Effects

The importance of addressing social and ethical concerns in the assessment of medical technologies has been noted in a previous OTA report (270) and will be discussed at greater length in chapter 4. Even though such concerns cannot ordinarily be valued, they are often essential to the measure of worth of a medical technology. The distribution of costs and benefits, respect for the autonomy of individuals, and a myriad of other social and ethical issues result from the introduction, extension, or modification of medical technologies (19). These issues are important to consider in assessments.

SOURCES OF DATA

All medical technology assessments require data. The chapters that follow discuss a variety of methods of assessment which can be used to systematically generate data about specific medical technologies and to produce information that is useful for setting policy. In addition, as described below, there are numerous systems which produce health-related data routinely.

The health statistics system of the United States is largely decentralized. Responsibility and authority for health statistical activities are divided among Federal, State, local, and independent agencies and organizations.

The compendium of statistical data sources in appendix A lists current public and private sources of data on the health of the population, the availability and use of health resources, and health care expenditures (especially as they relate to the assessment of medical technologies). Information for that compendium was obtained from a diverse group of individuals, governmental agencies, and nongovernmental organizations through data files, published reports, and personal interviews. Since each sponsoring agency or organization collects data using its own methods and procedures, the health data described in the compendium vary considerably with respect to source, method of collection, definitions, and reference period (93,282).

The largest single participant in the U.S. health statistical system is the Federal Government. The only Federal agency established specifically to collect and disseminate data on the health of the American people is the National Center for Health

Statistics (NCHS). Since 1960, NCHS has played a major role in the development of national health statistics policies and programs. The NCHS Division of Vital Statistics collects natality, mortality, marriage, and divorce statistics from the individual States and registration areas. In addition, NCHS conducts several general purpose surveys that provide statistics about the health status of the U.S. population. NCHS also has the primary administrative responsibility for the Cooperative Health Statistics System, a joint Federal, State, and local program for the collection of health data.[1]

An OTA study on Federal health data collection systems in 1979 found that the system for collecting, storing, processing, and disseminating health care information had been affected by the rapid growth of the Federal role in health care (282). At the time of that study, the Public Health Service alone administered 153 individual data projects; the Health Care Financing Administration operated at least 13 large statistical projects; and several other agencies and departments outside the Department of Health and Human Services (DHHS)* also conducted major health statistical activities. OTA concluded that these numerous data bases were uncoordinated and in many cases duplicative (282).

Four types of medical data sources are discussed further below: 1) data banks, 2) vital statistics, 3) insurance claims records, and 4) surveys of patterns of medical practice (197). Not every medical data system necessarily fits exclusively into one or another of these categories, but this breakdown of categories is useful for evaluating general strengths and weaknesses of different existing collections of data.

Data Banks

A potential source of information about patients, their characteristics, and their responses to the use of different biomedical technologies during their care are medical data banks.* Medical data banks, which are often computer based, are usually created by establishing a common terminology or vocabulary to describe a patient's clinical history and then entering observations on patients as events occur; sometimes, a data base management system is used to ensure accuracy, security, and easy entry and retrieval of observations. By extension, a medical data bank network contains the clinical history of large numbers of patients (from multiple centers) described in a uniform manner (218).

A medical data bank that contains extensive, comprehensive, reliable, longitudinal data on a number of patients can provide two important functions in assessing medical technologies. First, it can serve as an instrument for collecting the baseline and followup data on patients who have been subjected to a treatment. Second, the patients on whom this data has been collected can function as their own historical control group, which allows the investigation of the health effects associated with the treatment (93). Data banks can also be used to provide physician practice profiles and to assess the relative values of diagnostic and therapeutic choices.

Medical data bank demonstration programs have been funded by at least two Federal agencies: the National Institutes of Health (NIH) and the National Center for Health Services Research (NCHSR). NCHSR has supported a chronic coronary artery disease data bank program at Duke University and a rheumatic disease data bank program at Stanford.

Vital Statistics

Vital statistics such as those compiled by NCHS are of potential use in analyzing the safety and efficacy of medical technologies. When observational data are used to draw inferences concerning the safety and efficacy of medical technologies, however, there is a need to protect against bias in selecting the sample of records from which inferences will be made (197).

---

[1] Additional information concerning NCHS data systems can be obtained from *Data Systems of the National Center for Health Statistics* (93).

*Then the Department of Health, Education, and Welfare (DHEW).

*Although the terms are sometimes used interchangeably, a medical data bank differs from a registry (e.g., that used for cancer tumors). In a registry, data typically are abstracted from a specific document (e.g., medical record or death certificate) using criteria that are applied retrospectively, rather than by adhering to a common terminology in the prospective collection of data (56,218). In practical terms, registries generally cover discrete political or geographic areas (267). In this chapter, registries are discussed under the heading "Surveys of Patterns of Medical Practice."

Vital statistics data can have an important function in the area of record linkage. The National Death Index (NDI) recently put into operation by NCHS is a case in point (267). Historically, because there was no integration of records for the country as a whole, no mechanism had existed at the national level to determine whether a person has died. The NDI is intended to serve that purpose.

The NDI is designed to provide medical and health researchers with probable fact of death, the death certificate number, and the location of the death certificate when supplied with a minimum set of identifiers (generally the person's name and social security number or date of birth). A researcher may then contact the registration area where the possible match has occurred to obtain the death certificate or the required information. The NDI will be of immediate use in ongoing long-term studies which include mortality. Beebe (24) has described this index as the most important recent advance in making vital statistics accessible to researchers (24). The National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program plans to use the NDI to determine deaths of all persons in the SEER registries. SEER often loses track of people who move out of SEER areas before they die. Use of NDI should reduce the number of people lost to followup by SEER and provide better information about survival.

Insurance Claims Records

Large data sets have been compiled by third-party payers in both the private and public sectors. Blue Cross/Blue Shield, for example, collects data on its subscribers—about one out of three Americans, or more than 80 million people—as well as on subscribers and participants in various other programs that it administers (including Medicare and Medicaid in some regions) (197).

On the Federal side, passage of the Medicare and Medicaid legislation in 1965 created the need to establish a national yet decentralized administrative mechanism to pay for services to beneficiaries of these programs and to gather statistics for managing the programs. Systems needed to pay those bills were designed to provide information (primarily determinations of patient eligibility, the completeness of the claim forms, the appropriateness of length of stay in hospitalizations, and information on charges) for use by fiscal intermediaries in making interim payments. Although none of the data are collected solely for statistical purposes, the resulting information is useful for that purpose (198).

Surveys of Patterns of Medical Practice

Disease-specific (or procedure-specific) registries, hospital discharge data systems, and Federal statistical surveys of selected medical practices are a fourth general source of data for evaluating medical technologies.

A recent and somewhat prominent example of a procedure-specific registry is the voluntary registry established by the National Heart, Lung, and Blood Institute for physicians using the new technique percutaneous transluminal coronary angioplasty (PTCA). (The PTCA registry is discussed further in app. E.) Because it has achieved fairly high physician participation, this registry permits the collection of data for evaluation of PTCA at an early stage of the technology's diffusion, and at a cost that may be less than that of a randomized clinical trial (197).

Several hospital discharge abstract systems emerged in the early 1950's to provide summary information abstracted from hospital medical records about patients and their episodes of illness in short-term general hospitals. Although they tended to be established by independent organizations, many were similar in their origins, system characteristics, available data items, and sponsorship (which was generally private, nongovernmental).

The earliest system, the Professional Activity Study, developed in 1953 with support from the W. K. Kellogg Foundation. This system is operated by the Commission on Professional and Hospital Activities (CPHA), a nonprofit corporation in Ann Arbor, Mich., sponsored by the American College of Physicians, the American College of Surgeons, the American Hospital Association, and the Southwestern Michigan Hospital Council. The purpose of the Professional Activity Study is to link medical and surgical procedure rates with in-hospital mortality. The CPHA sample is a large one (approximately one-fourth of all hospitalized patients and one-third of the hospital discharges in the United States), and although not a random sample (it is based only on those hospitals that subscribe to their service), it has historically been representative of important hospital characteristics.

There are also an additional estimated 18 to 20 private, nonprofit hospital discharge abstract systems throughout the United States that process hospital utilization data for about half the hospitals in the country, representing over 20 million discharges yearly (198).

The Federal counterpart to these private, nongovernmental systems is the National Hospital Discharge Survey initiated in 1964. Administered by NCHS, this survey collects information on the characteristics of patients, lengths of stay, diagnoses, surgical operations, and the patterns of use and care. Only short-stay hospitals with six or more beds and with an average length of stay for all patients of less than 30 days are included in the sample (267).

PUBLICATIONS AND BIBLIOGRAPHIC DATA BASES

Information concerning medical technologies is often found in primary publications such as journals, books, Government reports, technical publications, and patents. Because of a dramatic rise in the number of publications in the field of medical technology assessment (13), however, access to the primary literature is often confusing and difficult. As a result, secondary publications—e.g., catalogs, indexes, bibliographies, and abstracts—which facilitate access to primary literature sources are increasingly important in the information transfer cycle. Secondary publications can also provide readers with superficial information about subject matter. (A bibliography on computed tomography, for example, can allow a reader to crudely approximate the state of the art solely by scanning the listed titles.)

Secondary publications are now increasingly available in the form of bibliographic data bases that can be read by a computer. The information contained in many of the bibliographic data bases (i.e., references to literature) also can be obtained "on-line." (A person at a computer terminal can carry on a dialog with a computer, directing it to locate, retrieve, and then display the information at the terminal or print the information on paper.) The growth of machine-readable bibliographic data bases in recent years has been extraordinary. In the United States alone, the number of data bases increased from 301 in 1976 to 528 in 1979, a 75-percent increase (394). Accompanying the growth in number of data bases has been a corresponding growth in use. The number of requests for information searches for individual data bases (on-line) grew from 700,000 in 1974 to an estimated 4 million in 1979 (395).

Medical technology assessment often may require information from many subject areas. Depending on the medical technology, the type of assessment, and a myriad of other factors, information may be sought in such diverse disciplines as biomedicine, law, finance, economics, and sociology. Such information is often obtained from computer-readable data bases. A common need in medical technology assessments is for information from the field of biomedicine. Information on biomedicine is found in numerous data bases. Worldwide, over 90 computerized data bases contain information on medicine alone (394). Although there is some overlap in the contents of many of the biomedical data bases, no one data base exactly duplicates another; each is unique in many aspects (e.g., contents, arrangement, and indexing philosophy). A descriptive list of biomedical-related bibliographic data bases of significance that are useful for medical technology assessment, along with the creator and vendor, appears in appendix B.

The major on-line service organizations, both public and private, that provide access to biomedical data bases in the United States are the National Library of Medicine (NLM), which is a part of NIH, Bibliographic Retrieval Service, DIALOG Information Service, Inc., and System Development Corp.

NLM provides access domestically and internationally, both to data bases in the biomedical field that are created and maintained solely by NLM and to data bases that it sponsors or produces in collaboration with other Government entities and private organizations, by means of its computerized bibliographic retrieval and technical processing system MEDLARS (Medical Literature Analysis and Retrieval System).[2] Its major biomedical data base is MEDLINE (MEDLARS On-line), which contains 600,000 references to biomedical journal articles published in the current and 2 preceding years. Tapes of MEDLINE and other MEDLARS' data bases, including TOXLINE* and HEALTH,** can also be leased from the National Technical Information Service (NTIS) of the U.S. Department of Commerce.

Bibliographic Retrieval Service, DIALOG Information Service, and System Development Corp. are commercial firms that do not produce data bases, but enter into licensing agreements with the data base producers which permit these firms to mount producers' data tapes on their own computer facilities, adapt the tapes to their own software (i.e., computer instructions), and sell on-line access to subscribers. These commercial vendors typically sell access to a broader group of data bases than just biomedical bibliographic ones. The commercial vendors sell access to MEDLINE and other MEDLARS data bases, as well as to health-related data bases such as EXCERPTA MEDICA and BIOSIS PREVIEWS which are produced in the private sector. They also sell access to non-health-related data bases.

LIBRARIES AND OTHER INFORMATION RESOURCE ORGANIZATIONS

The increase in the number and diversity of information products and services has been accompanied by an increase in the diversity of organizations that provide access to them.

Among the more important of these are health science libraries, which have for many years acquired, organized, and provided literature on biomedically related areas. As of 1979, over 2,700 public and private health science libraries were identified in the United States (74). The libraries were sponsored by medical schools, professional and vocational schools, business and industrial organizations, research organizations, societies and foundations, hospitals, area health education centers, health maintenance organizations, and health planning organizations (74).

In addition to having access to journals, books, and other print and nonprint materials, many health science libraries have access to biomedical and other computerized data bases. For example, about 1,500 health science libraries have terminals connecting them directly to NLM's MEDLARS system for computerized searches.* In addition, many other health science libraries provide indirect access to MEDLARS by referring requests to facilities with terminals. Indeed, informal and formal networks and the use of telephones, computers, and interlibrary loans have broadened access to information resource organizations considerably.

Although there are other types of information resource organizations which have provided, and can provide, information that is relevant to medical technology assessments, it is difficult to get comprehensive information about them—and even their typology is elusive and fluctuating.

---

[2] This system is the topic of an OTA technical memorandum entitled *MEDLARS and Health Information Policy* (276).

*Toxicology Information On-line. This is a collection of about 600,000 references from the last 6 years on published human and animal toxicity studies, effects of environmental chemicals and pollutants, and adverse drug reactions.

**Health Planning and Administration. This data base contains about 200,000 references to literature on health planning, organization, financing, management, manpower, and related subjects.

*The number and types of institutions which obtain access to the medical and other literature through the major commercial retrieval services (i.e., Bibliographic Retrieval Service, DIALOG Information Service, and System Development Corp.) are not known, because the information is proprietary.

For example, the first conclusion of a 1980 National Science Foundation (NSF) study and survey is (138):

> . . . that it is difficult to obtain comprehensive information on Federal S&T [science and technology] information centers. Even the National Referral Center at the Library of Congress is not provided accurate, timely, and complete information on federally supported information centers.

NSF restricted its definition of an information center in the 1980 survey to an organization whose primary function was:

> . . . to store, retrieve and distribute scientific and technical information. Both textual and numeric data information of a primary (e.g., information analysis) nature were included. Organizations such as agency libraries, which were involved in serving parent organizations, were in general excluded.

NSF identified 55 "science and technology information centers" in DHHS (then DHEW), including NLM.

A more recent study prepared for DHHS identified and categorized 157 "human resource information organizations" (6). Forty-one had a health focus. Of the 157 organizations, 98 were funded by Federal agencies, 43 by private organizations, and the others by academic institutions or State and local governments. The majority of federally sponsored organizations were funded by DHHS, but others were supported by the Departments of Education, Transportation, Housing and Urban Development, Energy, Labor, Justice, Commerce, Agriculture, and the Community Services Administration. Of the 157 organizations, 72 were classified as "clearinghouses," 76 were classified as "other types of information resource organizations," and 9 could not be classified because of incomplete information. (The other types of information resource organizations included special libraries, document depositories, information analysis centers, information referral centers, resource centers and networks, and technical assistance centers.)

Clearinghouses, which were the focus of the study, were distinguished on the basis of particular activities and functions. Through a variety of user services, these organizations perform three important tasks: the collection, the analysis, and the dissemination of information. They identify, select, acquire, process, and sort documents and other materials, and provide "locator tools" (e.g., indexes) to this collection. They also synthesize and digest this information to guide users to the specific data in the collection that best serve their needs. A wide range of clearinghouse services, from bulletins and announcements to bibliographies and handbooks, can lead users to the information they seek. Some clearinghouses tend to collect unpublished Government reports, projects, descriptions, speeches, and other types of "fugitive" literature that is hard to get elsewhere.

For example, the National Health Planning Information Center, a clearinghouse in the Bureau of Health Planning of the Health Resources Administration, was created to provide information for the analysis of issues and problems related to health planning and resources development. This center acquires, screens, and stores information on published journal articles, books, and other documents about health planning and resources development. Besides published reports, the center seeks unpublished reports, conference and proceedings papers, bibliographies, publication lists, and appropriate audiovisuals and microfilms. Its collection currently includes some 20,000 documents on a wide range of general subjects pertaining to health planning (e.g., health care technology and equipment impact, health care utilization). To facilitate the dissemination of information to health planners, the center issues selected publications in three series: *Health Planning Methods and Technology, Health Planning Information,* and the *Health Planning Bibliographies* (92,123,124). Finally, the center is now collaborating with NLM to include the center's data base in MEDLARS' HEALTH data base (276).

A second information clearinghouse is Project Share, which was created by DHHS to provide a central, systematic source of information for improving the management of human services. Targeted primarily at State and local officials, Project Share acquires, announces, and makes available documents relevant to the planning, management, and delivery of human services. Types of information collected and disseminated include published and unpublished papers, theses, research reports, bibliographies, technical reports, operating manuals, and conference proceedings. Besides providing a source of documentary and reference services, the clearinghouse analyzes and synthesizes reports and other documents describing human services program activities, conducts original research, and publishes state-of-the-art literature and state-of-knowledge reports (121).

The central repository for scientific and technical information generated by federally funded research and demonstration projects is NTIS. The NTIS collection exceeds 1.2 million titles. Most are drawn from the Departments of Energy and Defense, and approximately 10 percent have come from DHHS—5,700 reports from DHHS in 1979. In addition, NTIS has working relationships for the computerized processing of documents with at least three entities within DHHS: the National Cancer Institute, Project Share Clearinghouse, and the Bureau of Health Planning within the Health Resources Administration. NTIS is categorized as a clearinghouse, but the size of its collection and its function as the permanent repository of Federal technical information documents set it apart from all other clearinghouses except NLM (84,85).

CONCLUSION

The information needs for the assessment of medical technologies are both broad and deep, requiring the involvement of diverse disciplines. As the discussion in this chapter indicates, there exist numerous resources that are useful for assessing medical technologies. An earlier OTA study noted that a primary weakness of health-related data sources is the lack of coordination between them (282). This problem persists.

The next three chapters consider systematic methods for gathering and synthesizing information concerning the health, economic, and social effects of medical technologies.

3.

Evaluating Health and Economic Effects

> Ignorance never settles a question.
>
> —*Disraeli*

Contents

Page

Introduction 33

Evaluating Health Effects 33

- Experimental Studies 34
  - Randomized Clinical Trials 34
- Observational Studies 35
  - Cohort Design 35
  - Case-Control Design 37
- Summary 38

Evaluating Economic Effects 39

- Opportunity Cost 39
- Marginal Valuation 39
- Joint Production Considerations 40
- R&D Costs 40
- Overhead Costs 40
- Cost v. Prices 40
- Discounting 41

Conclusion 41

3.
Evaluating Health and Economic Effects

INTRODUCTION

Evaluating health and economic effects of medical technologies is central to any assessment; indeed, some would argue that evaluations of health and economic effects are the essence of an assessment.

The first section of this chapter discusses health effects. The main technical decision to be made when testing for the health effects of a technology is which study design is most appropriate. This section describes the study designs available and compares the designs presented in terms of their validity. Additional material on methods used to evaluate health effects is presented in appendix C.

The second section of this chapter concerns economic effects. It is primarily drawn from portions of a previous OTA report entitled *The Implications of Cost-Effectiveness Analysis of Medical Technology* (270). This section provides the reader with a brief discussion of the issues involved when evaluating economic effects.

EVALUATING HEALTH EFFECTS

Despite recent attention to the economic and social impacts of medical technology, the most critical aspect of the use of medical technologies remains their effect on health. An evaluation of health effects may examine efficacy (or effectiveness), safety, or both. *Efficacy** is the health benefit as measured under controlled conditions (such as those existing in a randomized clinical trial). *Effectiveness* is the benefit of technology under average conditions of use. Efficacy or effectiveness generally measure the intended effects of the use of a technology. *Safety* is a judgment of the acceptability of the risk** involved in using a technology.

There are many similarities between efficacy and safety—e.g., both are relative concepts and thus are discussed in terms of probabilities. Very importantly, however, their measurement may require different study methods. They differ in several key factors. In assessing efficacy, a study is usually oriented to a limited number of specific benefits. The measurement of safety, however, usually involves a study design that is able to identify a broad range of risks; such risks are often unknown or unexpected, they may occur far in the future, and they may affect only a small percentage of individuals. These factors imply that efficacy and safety are not simply the plus and minus columns of a single measure. Each requires separate attention, although judgments of the importance of either a benefit or a risk should only be made in relation to the other.

There are methodologic principles that guide the design, conduct, and interpretation of any particular investigation. Specific methods for evaluating health effects of technologies are described below. Each method has its strengths, weaknesses, and limitations for detecting favorable or unfavorable outcomes associated with a technology.

Of particular concern in research design and analysis is the validity of the findings, which varies with the study design. Validity refers to the extent to which a situation (as observed or evaluated by other criteria) is reflective of the "true situation." Four components of validity have been described (68). *Internal validity* refers to whether the observed effects of a medical technology, under the conditions of the study, are attributable to the technology and not to some other factors.

---

*OTA defines "efficacy" as "the probability of benefit to individuals in a defined population from a medical technology applied for a given medical problem under ideal conditions of use" (266).

**Risk is a measure of the probability and severity of harm to the health of individuals in a defined population associated with use of a medical technology applied for a given medical problem under specified conditions of use (266).

Statistical conclusion validity, which is a subset of internal validity, refers to the appropriateness of statistical tests and their ability (or power) to determine whether observed effects could be explained by chance fluctuations and to detect true differences in performance of the technology under study. *External validity* refers to the generalizability of the observed effects to other patient populations, settings, or conditions. *Construct validity* refers to the adequacy of the theory that an investigator has about what makes the technology effective.

The appropriateness of a particular study design is dependent on the purpose of the study, the methods available, the effects to be measured, and the technology's pattern of use. The choice of method is also influenced by other factors such as ethical concerns, limits on the number of participants available for study, the need for timely results, and available budget.

The discussion that follows is focused on select study designs which are commonly used to assess the health outcomes of a technology.

Experimental Studies

Experimental studies are characterized by the intentional application of a technology to a study population, and subsequent observation of effects. These studies must be carried out prospectively. They are frequently used prior to the dissemination of a technology, but can also be employed after the technology has diffused.

Randomized Clinical Trials

Randomized clinical trials (RCTs) are considered the most definitive experimental method for evaluating the efficacy or health benefits of a technology (60,148,187). An essential element of an RCT is randomization. Patients in an RCT are randomly assigned to one of at least two groups: one or more study groups, in which subjects are exposed to the experimental treatments, and a comparison group, in which the subjects are exposed to the control condition. The control condition can be either no treatment, the standard treatment (for comparison with a new treatment), or a variation (e.g., a different dosage) of the experimental treatment. The basic question to be answered in an RCT is: Are the effects observed in the experimental group also observed in the comparison group? If the answer is essentially "no," the effects observed in the experimental group can be attributed, within the limits of probability, to the treatment technology.

RCTs are a family of designs that vary in size and complexity. The number of treatment conditions (e.g., dosage levels) can vary, as can the size of population tested and the statistical power* of the study. Small RCTs may be performed early in the development of a technology to demonstrate or test the efficacy of the technology's innovative elements. Large-scale, multicenter trials can be conducted at a later stage in the development of a technology to establish its efficacy and safety across a large population and in diverse settings (266), as well as to increase the statistical power that results from a larger sample size. A major goal of the multicenter trial is to improve external validity in regard to larger populations.

Sometimes a favorable or unfavorable outcome is observed (i.e., a participant gets better or worse) because the participant *believes* that the treatment will work or believes the treatment is harmful. This "placebo" effect,** psychologically related but nonetheless real, results in a change in the participant's condition. Further, the effect may be influenced by the investigator's expectations. To reduce potential bias from the placebo effect, treatment can be offered under conditions where the participant ("blinding") or both the participant and the health care provider ("double blinding") are not aware whether the participant is given the experimental or the control treatment. Another layer of "blinding" is added when the person analyzing the data is not told which group is the experimental and which is the comparison. That person may be a statistician, but frequently is a medical specialist, and also may be the provider.

The principal advantage of RCTs is that they have high internal validity, i.e., they permit relatively unambiguous conclusions as to whether the observed effects of a treatment under the conditions of the study are due to the technology or some other factor(s). Randomization protects against potential selection bias in assignment of subjects to experimental and comparison groups. Within the limits of sampling error, the only difference between the groups is that the experimental group is given the treatment under study and the comparison group is not. Therefore, differences in outcome can be attributed to the differences in treatment, with a known probability of error due to chance.

---

*The "power" of a study is the probability of its detecting an effect (of technology being tested) when one actually exists. The greater the power, the less likely one is to incorrectly reject an effective technology.

**Although the placebo effect is discussed here under RCTs, it is not peculiar to such studies.

Although well-designed RCTs are generally high in internal validity, they do not *necessarily* resolve the problem of external validity (68). External validity is usually established only when large heterogeneous samples of participants are tested under a variety of circumstances, typically across a number of studies, or through large multicenter RCTs with carefully selected populations.

A disadvantage of RCTs is that they can be difficult to carry out in settings such as hospital clinics and physicians' offices and can be especially difficult for technologies that are already widely diffused and perceived as being effective (253, 401). In such situations, administrators and clinicians may be reluctant to make the changes in policies and procedures needed to conduct an RCT. Preexisting conclusions on the treatment being evaluated are a major obstacle to conducting RCTs (159). Such conclusions may subvert the randomization process. For example, the assessment of high-oxygen environments as a cause of blindness in premature infants was impeded by well-intentioned nurses (346). In one study nurses raised the oxygen level for the experimental group of infants in the belief that the low-oxygen environments were harmful. In another study, it was necessary to implement the treatment only partially, until evidence of the harmful effects of oxygen were more apparent.

RCTs are generally considered more complex and expensive to conduct than other types of studies. The decision to initiate an RCT should be based on strong evidence that the hypothesis under consideration merits the possible expense and effort of conducting such a study.

Finally, RCTs may be of limited utility in studying safety. As indicated above, safety is a measure of risk, and risks may occur after a considerable time, may occur infrequently, and may be unexpected. These types of effects may be difficult to plan for and measure by an experimental study, thus necessitating the consideration of other forms of assessment.

Observational Studies

Observational studies may be valuable in generating or testing hypotheses about the health effects of a technology once the technology is widely diffused. They also may be considered in situations where experimental studies are inappropriate or impossible to conduct. The common element in all observational study designs is that the investigator does not control the application of the technology under study. The division of a population group into "cases" and "controls" or "exposed" and "unexposed" occurs through mechanisms unrelated to carrying out a study, such as the treatment preference of a physician (e.g., in the care of a stroke victim) or self-determination (e.g., in the choice of a method of contraception). Although the internal validity of observational study designs generally does not match that of experimental study designs, observational studies may allow evaluators to rule out competing explanations for the observed effects.

Because the investigator does not employ the deliberate or intentional modification of conditions between the study groups, steps must be taken to try to eliminate any potential bias in selecting the study groups. The investigator must try to control for bias, which may result when groups differ with respect to "confounding variables" (age, sex, health status, or any other characteristic which may account for observed outcomes). However, in nonrandomized studies the extent of selection bias cannot be known, and thus the effectiveness of the steps taken to minimize bias also cannot be known with certainty.

Cohort Design

Cohort studies begin with a "naturally occurring" population, or a sample thereof, chosen by the investigator as defined by: 1) some criterion or combination of criteria such as specific age, location, time period, etc., and 2) exposure or nonexposure to a technology. The population is followed over time to observe the differences in health status between the exposed and unexposed groups. In a "prospective cohort study," the population is identified at the time of exposure and health status is assessed at a future time. If the population is identified after the exposure has occurred and the health status of the individuals is assessed at the present or a future time, it is termed a "retrospective cohort study."

A 1978 study by Roos and colleagues (319) employs a retrospective cohort study in assessing the effectiveness of tonsillectomy (with or without adenoidectomy) in preventing subsequent episodes of respiratory illness. This study illustrates many of the features that can be built into cohort designs to minimize the effects of confounding variables in an attempt to improve internal validity.

Roos and colleagues used medical claims and patient registration data provided by the Manitoba Health Services Commission to identify the population from which the cohorts were drawn. Two operated groups were created: one consisting of all patients operated on for tonsillectomy only during January 1973; the other consisting of all patients operated on for tonsillectomy only or tonsillectomy plus adenoidectomy for all of 1973. In addition, two comparison (nonoperated) groups were formed: children under the age of 14 whose records indicated evidence of tonsillar illness but no tonsillectomy operation during a 3-year period (1972-74); the other consisting of nonoperated siblings of operated patients.

Analysis of the data indicated that, on the average, the surgical procedures averted about one episode of respiratory illness per child over the 2 years following surgery. The greatest benefit accrued to the patients who had experienced the greatest number of episodes of respiratory illness in the year preceding surgery.

Had the investigators not taken measures to control for confounding variables, the observed results might have been explained by factors other than the surgical procedures. Specifically, maturation (changes in health with age) was not believed to affect the results since the groups were similar in age and sex. The effects of "history," temporal events such as new health practices, was minimized by the use of concurrent controls. "Local history," the effect of family preference or physician factors, such as a predisposition toward surgery, was also minimized by including the sibling control group. These design features presumably minimized threats to the internal validity of the study.

Postmarketing surveillance, the mechanism used to detect unsuspected adverse drug reactions *after* a drug is marketed, generally employs the prospective cohort design.* Typically, a user population of a particular drug is entered into a registry and followed over time for various "health events." Rates of such events are compared with rates in a nonuser population. Thus, unusual medical events may be associated with use of the drug. These studies, because they use relatively large populations, may detect associations between drug use and unusual adverse reactions which are generally not detected in small population studies such as those used in premarketing assessment of the drug.

Historical Controls.—Innovations in medicine often diffuse so rapidly and completely that new technologies or new treatment variations may become standard in a fairly short time. It may be impossible to assess the long-term outcome of the technology in a conventional prospective cohort study, for lack of a group of patients not given the new treatment. In these instances, if researchers are to conduct a study, they may use a variant of the cohort design which employs historical control groups, i.e., patients treated prior to the innovation. The use of historical controls, however, adds a serious limitation to the cohort study design: change, other than the change in the treatment being assessed, is constantly occurring in health care, and such change may affect the internal and construct validity of the study. Yet despite their limitations, historical control studies can be useful, particularly if the temporal gap between control and treated groups is small, since

---

*For a detailed discussion of postmarketing surveillance, see OTA's report entitled *Postmarketing Surveillance of Prescription Drugs* (281).

the likelihood of some validity problems is reduced. Great care, however, should be exercised in their use and interpretation.

One of the first studies assessing the efficacy of coronary care units (CCUs) (314) relied on historical controls. The first 200 patients with acute myocardial infarction (heart attack) admitted to the CCU at Royal Perth Hospital formed one cohort, and the last 200 patients treated for acute myocardial infarction prior to the opening of the CCU formed the comparison group. Although mortality rates by severity of infarction were somewhat better for patients treated in the CCU than for patients in the historical comparison group, the difference was not statistically significant. Because the patients were all treated at the same hospital (though at different times), the two groups were similar in many respects: the base population was similar, the hospital staff was basically the same, and hospital records, on which the study relied, were similarly kept. The validity of the study (i.e., did the CCU produce the effect), however, was compromised by the introduction at about the same time as the CCU of a number of other therapeutic measures (e.g., lidocaine and atropine to treat and prevent arrhythmias and the use of transvenous pacemakers to treat conduction blocks). This study of CCU efficacy was not definitive, and the value of CCUs, themselves not strictly defined entities, is still an open question. However, the study did raise enough questions to spawn further investigations.

Studies of the use of high-dose methotrexate chemotherapy for treating osteosarcoma, a form of bone cancer,* illustrate a case where the use of historical controls so compromised the study that erroneous results were obtained. Following the development of chemotherapy in the early 1970's, researchers began to experiment with ways to improve its apparent effectiveness. One approach was to treat patients with drugs before their cancer had spread. Studies using historical controls indicated that nearly half the patients treated in 1970 lived 2 years without a recurrence of the disease, compared to only 20 percent of a group of patients treated in 1960. However, the change in therapy from 1960 to 1970 was accompanied by other changes in diagnosis, treatment, and patients. For example, the patient mix undoubtedly changed over the 10-year span so that patients with the worst prognoses (i.e., metastatic cancer) no longer constituted the majority of those treated, rendering the cohorts noncomparable. One can have little confidence in the results of a study which seems to show the chemotherapy efficacious, when the confounding effects of the other secular changes that occurred between 1960 and 1970 could account for the effects of the treatment in analyzing the study data. In particular, the Mayo Clinic found that patients *not* treated with chemotherapy in the later time period also had higher survival rates.

In summary, cohort studies using historical controls serve a limited but sometimes helpful purpose. They may allow for an inexpensive preliminary inquiry as to the value of a technology, capitalizing on existing data. However, they seldom, if ever, provide definitive information on which to make decisions about the value of the technology.

Case-Control Design

Case-control studies compare a group of people with a disease (or other outcome event), cases, to another group without the disease, controls, and then determine whether they differ in their previous exposure to a presumed causal agent (e.g., a drug). These studies are retrospective in nature, the exposure having occurred prior to the identification of cases and controls.

Substantial biases are possible in case-control studies. The most serious result from the selection of an inappropriate comparison (control) group. Because it is not possible to achieve complete comparability between the comparison group and the case group, controversies about the interpretation of case-control studies generally revolve around the question of whether or not the controls are an appropriate representation of the population that gave rise to the cases. Other problems also exist: the retrospective nature of the method implies no control over the treatment, forces reliance on individuals accurately recalling past events, and forces reliance on records that were kept for reasons other than those of carrying out a study.

---

*G. B. Kolata, "Dilemma in Cancer Treatment," *Science* 209:792, 1980.

The studies of a possible association of estrogen therapy with endometrial cancer illustrate the problems encountered in using case-control designs. The major dispute among researchers (191, 196) concerns the appropriateness of the control group. The traditional approach, to compare patients with endometrial cancer to control patients with other genitourinary cancers, has found a consistently high association between the use of estrogen and endometrial cancer. Critics of this approach note that because estrogen use may provoke uterine bleeding, and because a woman with bleeding is very likely to seek medical attention, there may be a higher percentage of women carefully examined and tested in the group taking estrogens than in the population of women not taking estrogens. This would lead to a higher rate of detection of endometrial cancer in the estrogen group than in the nonestrogen group.

Horwitz and Feinstain (191) contend that because of this increased surveillance, cancers are detected in the estrogen group that otherwise would not come to clinical attention during the lifetime of the women, and that if the nonestrogen group were tested as carefully, more cancers would be detected in that group. To counteract this potential selection bias, these investigators recommended selecting controls from among women surgically treated for noncancerous uterine diseases. The use of such a population to create the control group should adjust for the bias resulting from increased surveillance and diagnosis. Horwitz and Feinstein showed that when this selection procedure was employed, the likelihood of estrogen being linked to cancer was significantly lower than under previous study approaches. As Cole (61) has stated, however, patients undergoing the same diagnostic procedures as the cases can be "an inappropriate control group," since the same causal agent may be responsible for their illnesses.

These studies have not resolved the issue, however, and proponents of traditional control selection procedures claim that there is little detection bias in their method, since most cases of endometrial cancer are eventually diagnosed (196). These critics maintain that the controls used by Horwitz and Feinstein are biased, because they do not give an appropriate picture of estrogen use in the underlying population. However, recent evidence from autopsy studies has shown that many cases of endometrial cancer indeed *are* unsuspected during life and are first detected, if at all, at autopsy (192).

Is there more or less bias in Horwitz's and Feinstein's control group selection than in the traditional approach? The two approaches might be viewed as providing a range of estimates for the relationship being examined. Because of the internal validity problems associated with this method, the use of different control groups to estimate the range of relative risk estimates might be considered.

In summary, case-control studies are relatively inexpensive, can be carried out in a relatively short time and usually employ smaller sample sizes than other study designs. The case-control design lends itself to ascertaining the associations between known rare events or outcomes and suspected causal agents when the events occur only years after the exposure. For example, this design might be used to investigate the relationship between a commonly used drug and a rare adverse effect. Case-control studies can be used to explore a hypothesis without disrupting medical practice. However, case-control studies are not useful for discovering previously unsuspected effects or discovering adverse effects of rarely used drugs.

Summary

Observational study designs used to assess the outcome of a medical technology are those in which the investigator does not control the application of the technology to the study population and applied in essentially the same manner as observational designs would be applied to examine other risk factors for disease. These designs are most applicable for detecting or ruling out specified but unforeseen adverse consequences of a technology after the technology has been diffused. Experimental designs, those in which the investigator controls the application of the technology according to specific criteria, are in theory and often in practice more useful, especially in determining efficacy.

The degree of validity, particularly internal and external validity, of the findings varies with the study design chosen. Observational studies (e.g., case-control and cohort) lack the high degree of internal validity found in the design of choice for experimental studies, the RCT. That is, it is usually more difficult in observational as opposed to experimental studies to determine whether the observed differences can be attributed to the technology under study. Because observational studies can more accurately reflect the conditions of use of medical technologies in the population, they may, in some cases, have a higher degree of external validity than experimental studies.

The study design ultimately selected depends on several factors, including the developmental stage of the technology, the purpose of the study, ethical considerations, the population available, and budget constraints. Seldom is assessment a one-time event. Associations of cause and effect can *rarely* be established through a single study. In theory, judicious decisions in study design selection should be based on a review of previous and ongoing studies so that each new study becomes a building block toward the total assessment, leading to sound policy decisions. In practice, sometimes they are not.

EVALUATING ECONOMIC EFFECTS

The economic effects of medical technology have been assessed through a variety of methods, most notably cost-benefit analysis, efficiency studies, cost-effectiveness analysis (CEA), cost-impact (total costs associated with a technology), or private sector-oriented techniques such as return-on-investment analysis.

Currently, the most visible and potentially the most useful of these techniques is CEA. CEA is not simply an economic technique; it is a blend of economics and clinical information. As such, it will be described in chapter 5 with synthesis.

No matter which form of analysis is chosen, certain methodologic considerations need to be taken into account. These considerations were identified and examined in previous OTA reports (270,271), and the discussion presented here is based on that earlier work.

Opportunity Cost

The principal concept when evaluating the economic effects of a medical technology (or any activity) is opportunity cost. The opportunity cost of a resource is its value in its next best use. Thus, the true cost of a resource is not necessarily its market price tag. Rather, it is what one must give up elsewhere in order to use that resource.

An illustration should help to clarify the difference between a market price tag and a resource's true opportunity cost. From the perspective of a hospital accountant, volunteers' time is free; it is not found on the hospital's wage bill and the accountant would ignore it. But is volunteer labor not a true cost of running the hospital? Volunteers definitely contribute to the output of the hospital. And from a social perspective, if the volunteers' labor would have been donated elsewhere had the individuals not worked at the hospital, such labor clearly has value. In essence, the opportunity of using the labor productively in other activities has been foregone. From a social perspective, therefore, the volunteers' time should be included in an assessment of costs. Although determining an appropriate dollar value may be difficult, the social value of volunteer time should not be ignored in an analysis.

Furthermore, as stated in chapter 2, both direct costs—resources purchased directly—and indirect costs—the value of the lost "production" time seeking care or being sick—should be included in an analysis.

Marginal Valuation

The worth of a technology should be assessed at, what economists term, "the margin." That is, an analysis should seek to compare the added, or marginal, cost of producing the next unit of benefit.

In an evaluation of computed tomography (CT) scanning, the issue is no longer whether the technology itself is cost effective, but, rather, whether the various applications of the technology are cost effective. Should CT be used for confirming suspected brain disease/trauma, or for ruling out brain disease/trauma when persistent headaches are presented? In what instances are body scans indicated—or cost effective?

In general, the relevant inputs or costs which must be considered in the case of a medical technology will be tied to whether the technology is already in place or whether it has yet to be adopted/purchased.

Joint Production Considerations

Many technologies have multiple applications, and the technological process being studied is seldom applied in isolation. These two considerations can have enormous effects on cost calculations.

For instance, since a single blood test can be and is often used as a source of information for numerous diseases and bodily functions, analyzing the cost of drawing blood for only one purpose is inadequate if the total cost is used; it either overstates the associated costs, understates the potential benefits, or both. Likewise, a CEA of a Pap smear program should be done in recognition of the fact that many other health evaluations are not only possible but are ordinarily performed during the examination, whether formally or informally. That is, a woman who is given a Pap test may be screened for other pelvic disorders, high blood pressure, fever, skin rashes, weight problems, and many other conditions. All of these procedures carry certain potential benefits and all of them should be assigned some of the cost (or, conversely, less cost should be assigned to the Pap test); or the analysis should be evaluating the complete examination rather than just the Pap test.

Including the effects of joint production adds greatly to the problems of measurement and valuation, but these difficulties in no way diminish the conceptual importance of fully considering these effects in a complete CEA. Sometimes, for instance, a very small incremental (or marginal), increase in cost to an existing production process can have large benefits spread over multiple applications. However, some large cost increases may produce fewer benefits than existing production processes when their contributions to all the applications are taken into account.

R&D Costs

R&D costs may pose a problem when evaluating a technology's worth. In general, where R&D is an integral part of the immediate program in question (e.g., when analyzing the costs and benefits of a new technology in a medical research center), the R&D resources should be included along with the program's operating inputs. When the R&D has preceded the program being evaluated—that is, its existence is independent of the immediate policy decision—R&D resources should be excluded from consideration.

Overhead Costs

Determining how to allocate overhead costs is particularly difficult. If the use of the technology at issue is truly marginal to the overall enterprise, one might be tempted to ignore overhead, to look only at the marginal resource needs associated with the program. However, if the existence of some of the overhead depends on the program in question, clearly it must be identified and included. The general principle of seeking the marginal inputs still holds, but often in practice one may have to attribute to the program a share of overhead proportional to the program's share of the total enterprise.

Costs v. Prices

Uncritical use of market prices can lead to large gaps between cost estimates and true costs. Illustrative of this problem is the use of hospital charge data to reflect the costs of hospital care. A common practice, this form of "pricing" ignores the known idiosyncrasies of hospital accounting in which hospitals charge well above true marginal costs for certain services and use the profits to subsidize other services for which charges do not cover marginal costs. If the deviations from marginal costs were small, one might reconcile acceptance of imperfect hospital data as a readily available source of information providing a qualitative valid picture. However, studies of the discrepancies between true costs and charges show dramatic differences. For example, hospital pharmacy charges can vary from 10 to 1,000 percent of the true costs of drugs depending on the frequency of their use, their level of cost, purpose, etc.

Discounting

Costs and benefits seldom occur at the same point in time. Through the application of a method termed discounting, however, they can be treated as if they all occurred in the present.

The rationale for discounting future costs and benefits stems from the fact that resources can be productively invested for future gains, as well as from the observation that people expect to be rewarded for postponing gratification. For instance, in order to induce individuals to save, interest must be paid, even in the absence of inflation. The rate of interest determines the future value of the amount invested. Thus, for example, $100 invested at 5-percent interest this year will become $105 next year. Discounting is the reverse process: $105 next year has a "present value" of $100 when the discount rate is 5 percent.

Although there is general agreement among economists and policymakers that discounting future moneys is conceptually correct, there is no consensus concerning what discount rate should be used, and there is still some confusion as to the proper method of valuing future nonmonetary benefits/effectiveness. However, when benefits are long delayed, almost any discount rate will reduce benefits substantially (to near zero in extreme cases), making them less important to the outcome of the analysis (270). Thus, this phenomenon results in making the rate used and the uncertainty of future events less important than they otherwise would be.

CONCLUSION

Choosing a research method for assessing health effects depends on various factors. In general, one should opt for the study design that produces results. But constraints such as economic and social/ethical factors limit one's choice. There is a role in medical technology assessment for each of the methods discussed in this chapter. The important point to remember is that each method has its inherent strengths and weaknesses, and one must always exercise caution in accepting the results of a study without carefully taking note of the study's limitations.

Evaluating economic effects requires careful attention to the principles outlined in the latter portion of this chapter. An important point to always keep in mind is that costs are usually not what they appear to be, especially in health care.

The evaluation of the social and ethical implications of medical technologies is discussed in the next chapter. The following chapter reviews methods for synthesizing results from research studies, CEA, and group decisionmaking techniques.

4.

Social Values in Technology Assessment

The very success of science has ended its pleasant isolation.

—*Robert Sinsheimer*

Contents

Page
Introduction 45
Value Premises of Medical Technology Assessment 45
Social Implications of Medical Technologies 47
  End-Stage Renal Disease Program 48
  Maternal Serum Alpha-Fetoprotein 48
  Artificial Heart 48
  General Comments 49
Policy-Related Activities for Considering Social Effects of Technologies 51
  Office of Technology Assessment 51
  National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 51
  Ethics Advisory Board 52
  President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research 51
  National Center for Health Care Technology 52
Conclusion 53

4.

Social Values in Technology Assessment

INTRODUCTION

Any decision to develop or use a medical technology, or not to do so, inevitably rests on value judgments, though such values may not be explicitly acknowledged. One goal of technology assessment, therefore, is (63):

> . . . the identification of [social] implications possibly overlooked by decisionmakers who, by their *own* values, would not want them overlooked or implications that decisionmakers cannot afford to ignore, even if they so desire, because many other people, by *their* own values, may find them important.

This chapter explores the role of values in medical technology assessment. The first section examines the value premises of the assessment process itself. Implicit value judgments permeate every stage of a technology assessment, and it is important to recognize these when conducting or examining a specific assessment. The second section of this chapter considers the role of value analysis in assessing the social and ethical implications of policy decisions regarding the development and use of medical technologies. The third section reviews the efforts of some past and present Federal bodies to consider these implications.

VALUE PREMISES OF MEDICAL TECHNOLOGY ASSESSMENT

Technology assessment and other types of policy analysis can never be totally objective or value-free. Even at the most fundamental level, technology assessment is based on the value assumptions that: 1) it is better for society to systematically analyze the far-reaching consequences of technological change and development for its effect on economic, social, and ethical values; and 2) it is better for society, in terms of maximizing benefits, minimizing risk, and promoting efficiency, to have the information that technology assessments can provide. That these things are "better" represents a fundamental assumption, an unspoken value judgment.

Indeed, although they are rarely made explicit, value judgments enter into every phase and aspect of technology assessment. Such judgments determine: 1) which technologies will be assessed and at what stage of their development, 2) what the scope of assessments will be, 3) what kinds of data will be collected and how these data will be analyzed, 4) what methods will be used in an assessment, and 5) how the results of an assessment will be presented and interpreted.

The first thing to consider, perhaps, is the genesis of the assessment process. Assessments of medical technologies are requested by decisionmakers concerned with questions relating to their safety and efficacy, reimbursement, and appropriateness. The factors and perceived needs that lead to a specific request should be clearly understood by those conducting or examining an assessment. The constraints inherent in the request should also be understood. What types of issues is the assessment not expected—or not permitted—to consider? Finally, the assessment should be put into a larger perspective. Why is the technology a concern to society? Where does the assessment topic fit into the current cultural and political setting?

Also important to consider are the values of the assessors and of the analytic methods, goals, and objectives they choose to employ. Here the issues are most numerous. First, there is a need to determine the boundaries and scope of the assessment. What values should the assessment include? If value judgments are organized into a hierarchy/continuum of abstractions, running from generalities to specifics, a technology assessment begins at some point in this hierarchy/continuum, with the implicit assumption being that only the value implications *below* that point will be considered. Values above that point (i.e., those considered more abstract or general) are prior assumptions that are accepted, laying a foundation for the analysis that follows. In other words, an agreed upon set of decisions is in some sense "final," at least from the perspective of the analysis at hand, so that the analysis is only concerned with specific policies or refinements (214).

In some sense, establishing a hierarchy of value judgments incorporates the values of the individuals performing the assessment. Nevertheless, the values of the assessors warrant special attention on their own. It is unreasonable to presume that one can begin to appreciate, in any worthwhile fashion, the psyches of those conducting an assessment. The concern, however, should not so much be what the assessors' individual biases are as with ensuring that their values do not overly bias the outcome of the assessment. Thus, it is important to build into the assessment measures to correct for possible value distortions. Bioethicists have been particularly successful in this regard, bringing a broader set of values—more representative, perhaps, of those held by the public—to the fore in technology assessment.* The key is to look for values that have been consciously or unconsciously omitted from the analysis. Attempts to broaden the values represented in an assessment can also be made by performing an assessment under public scrutiny. In conducting its assessments, for example, OTA uses multidisciplinary advisory panels that include interested parties.

The "tools" of technology assessment—e.g., the methods of collecting and evaluating data—must be applied with great caution and with the broadest possible understanding of the kinds of distortions they can create. One danger in technology assessment is that problems may be reduced to terms that mistake their underlying structure and ignore their total character. Indeed, the problem of using too narrowly defined objectives is of concern in all policy analysis. Convenience for the analyst often leads to inaccurate definitions of problems.

The measurement and analysis of data are tasks involving more than technical procedures, and carry implicit value systems and orientations (140). In assembling data and information relevant to an assessment subject, value judgments are incorporated in the choices of what to look for, the manner in which data are measured, and the manner in which information is presented.

For example, in measuring improvements in health status, changes are frequently expressed in levels of resource inputs. It is not clear, however, that inputs such as more doctors, more hospital beds, more computed tomography (CT) scanners directly translate into improved health status. Furthermore, even if one assumes that *more* of a resource input is "better," one still cannot measure the improvement in health status that results from the addition of each resource unit. If decisionmakers are attempting to determine which resource is needed more, they must know the levels of effectiveness involved in the addition of doctors, beds, or CT scanners. Otherwise, there is no basis for choosing among programs emphasizing one of these resources. In addition, measuring health status in this fashion carries the implication that the development of any new programs must be biased in favor of such measurable goals, or the new programs cannot be compared against older programs (140).

The methods available to assess medical technologies also have normative underpinnings. If, upon examination, the underlying normative assumptions are found to be unsatisfactory, the conclusion of the analysis must be rejected. OTA's report on cost-effectiveness analysis (270) discussed in detail the inappropriateness of aggregating costs and benefits in economic analyses. Thus, for example, the technique of cost-benefit analysis is of concern, because it requires aggregation which tends to ignore the distributional effects of the costs and benefits that the technique attempts to measure. Encouraging the conduct of randomized clinical trials implicitly signals a willingness to subject a relatively small number of individuals to varying degrees of risk, hoping that greater benefits will accrue to society as a whole (19). The danger here is that one small group (e.g., the urban poor) may bear a disproportionate share of the risks of experimentation, raising serious questions of equity and autonomy.

---

*Bioethics is a discipline that brings analytic rigor to considering values camouflaged or implicit in medical, biomedical, and health care issues. In conducting assessments, one may wish to involve bioethicists and others with expertise and experience in moral or ethical principles to identify and analyze relevant moral principles in available policy options. Though experts in value analysis may be quite proficient in arraying ethical and social implications, this does not imply that their expertise qualifies them to select which of those principles society should pursue most vigorously (377).

Thus, throughout an assessment there is a need for constant inquiry into the nature of the questions being asked. Despite deficiencies in measurement techniques and the difficulty of translating social principles and values into practical terms, technology assessment can contribute to better judgments regarding appropriate responses to technological change and development when social values are explicitly considered.

SOCIAL IMPLICATIONS OF MEDICAL TECHNOLOGIES

To varying degrees, medical technologies may directly or indirectly influence the quality of the lives of individual patients and their families; the structure of medical, legal, political, and economic systems; and the fundamental values on which these social systems rest, including society's sense of ethics and morality.* These "social implications" cannot easily be quantified, but they can be identified and rigorously analyzed.

An important task in medical technology assessment is to identify the conflicting social values that underlie policy alternatives. This assessment task is sometimes referred to as "value analysis."** In the broadest sense, the task of value analysis in examining social implications is to bring into focus the compromises that are made with society's preexisting goals, values, and institutions when choices are made between policy alternatives. Often, value analysis may simply provide a more conceptually clear understanding of policy problems by describing the complex interaction of interests within the confines of established social—economic, medical, legal, and cultural—values.

Some work has been done in the area of suggesting techniques for assessing social implications. In 1976, OTA developed an illustrative list of questions that could be asked regarding a medical technology (269):

- What are the implications of the technology for the patient?
- What are the implications for the patient's family?
- What are the implications for society?
- What are the implications for the medical care system?
- What are the implications for the legal and political systems?
- What are the implications for the economic system?

Wolf has suggested that certain economic analytic techniques (e.g., cost-benefit analysis) could be modified and applied as value analysis (397).

Jonsen and Butler have suggested that value analysis specifically in the area of ethical implications should be structured by the following three tasks (205):

1. articulation of relevant moral principles;
2. elucidation of proposed policy options in light of the identified principles; and
3. rank ordering of policy options for choice.

The first task is to identify the moral and ethical principles around which the policy issue turns, and to set these principles into the center of the discussion in as definite form as possible. In setting public policy to guide the conduct of biomedical research on fetuses, for example, two apparently conflicting concerns overwhelm the debate: respect for the autonomy of individuals v. the knowledge (and hence the benefits) society gains through such research efforts. The second task is to examine how policy options interact with the various identified ethical principles and theses. This task involves identifying and isolating the subtleties of ethical questions. Finally, the third task is to rank the policy options to show how each policy would look and what its probable outcome might be if one moral principle were ranked over another. This ranking is similar to economists' arraying of tradeoffs in costs and benefits when comparing alternative policies (270,383).

---

*Ethics comprises "the principles of morality, of right and wrong conduct, of virtue and vice, and of good and evil as they relate to conduct" (304). "Ethics" and "morality" are sometimes used interchangeably.

**Value analysis and the social implications of medical technology are most often discussed by bioethicists. Much of the bioethics literature focuses on the ethical implications of medical technologies, as ethical issues are often the most profound and difficult to reconcile. However, not all of the social implications of medical technologies are ethical ones.

Public policies in the United States are not directed toward a single set of objectives. Different policies reflect different social goals, which may often appear to be in conflict. In determining whether to further the development or use of a medical technology, questions of safety, efficacy, and cost effectiveness are centrally important, but in some cases—especially when the technology brings two or more of society's values and goals into conflict—these may be outweighed by a broader set of costs and consequences. Such conflicts are illustrated by the technologies discussed below.

End-Stage Renal Disease Program

This year, the Federal Government will spend over $1.2 billion to provide dialysis treatment and kidney transplants to some 50,000 patients suffering from end-stage renal disease (ESRD) (see case study in app. E). Without such treatment, these patients would die from renal failure. Although no one denies that the Federal ESRD program provides medically necessary services to people in dire need, the program is surrounded by agonizing questions for policymakers faced with decisions about allocating medical resources. Because of their enormous costs, disease-specific programs (like that for ESRD) cannot be publicly funded for *all* patients whose medical needs are equally pressing or more ambiguous (48).

The ESRD program was enacted as a humanitarian response to the vivid impact of dialysis and transplants and the plight of needy patients. In that it does not offer guidance for selecting among equally needy groups of patients suffering from other diseases, the enactment of this program does not reflect a guiding ethical principle. The absence of analytic foresight in this instance makes the ESRD program appear to be a political accident. Because choices among programs competing for scarce resources inevitably do and should rest on value judgments, more coherent public policies might evolve if attendant analyses represent a careful working through of the ethical underpinnings of policy alternatives.

Maternal Serum Alpha-Fetoprotein

Maternal serum alpha-fetoprotein (MSAFP) (see case study in app. E) is the first in a series of diagnostic tests used to screen and diagnose two types of fetal neural tube defects: anencephaly (absent or undeveloped brain) and open spina bifida (failure of the spine and overlying skin to fully close over the spinal cord). Initially, attention was focused on MSAFP because of profound social implications inherent in its diffusion and use. The test is given to expectant mothers to provide them with information about the fetus, the value assumption being that it is better for them to know in advance if their children are to be born with neural tube defects. Since some mothers given information that such defects are present might be expected to terminate their pregnancies voluntarily, MSAFP was thrust into the ethical argument over abortion.

The test has also raised questions of distributive justice, particularly with regard to entitlement programs. For example, the Health Care Financing Administration was concerned with the ethical implications of reimbursable MSAFP tests for Medicaid recipients. Women receiving Medicaid were not entitled to abortions. Thus, the dilemma arose: If Medicaid agrees to reimburse physicians for providing MSAFP tests, what happens to women with test results indicating fetal defects when abortions are not reimbursable?

Artificial Heart

Another medical technology that illustrates the importance of social implications is the artificial heart. In 1972, the National Heart and Lung Institute* convened a panel of physicians, ethicists, lawyers, and social scientists to identify and evaluate the "economic, ethical, legal, medical, psychiatric, and social implications of a totally implantable artificial heart." The institute was concerned about the broader implications of such an innovation, implications which the institute's physicians and administrators recognized were "beyond the limits of their own expertise" (204).

The panel primarily focused on two sets of questions, both related to the ethical problem of distributive justice. First were questions concerning access to the device and the selection of patients to receive what would most likely be a scarce, expensive medical resource—problems endemic to modern medicine. Second were questions applicable only to an artificial heart powered by nuclear energy. (The National Heart and Lung Institute had been developing three power systems for its device: an electric motor powered by a biological fuel cell, a motor powered by rechargeable batteries, and a nuclear engine fueled with plutonium.) The panel noted that patients faced with imminent death from heart disease might be willing to accept the attendant risks of prolonged exposure to radiation from the nuclear device, but it expressed greater concern for persons exposed to "slight though significant risks" through contact with recipients. A majority of the panel's members doubted that a decision to make a nuclear-powered artificial heart widely available would be "ethically justified when measures to improve the health and extend the life of specific individuals pose a risk to the health and lives of the population generally, including unborn future generations" (204).

General Comments

Not every new medical technology warrants a full-scale technology assessment with an examination of social implications. Some technologies probably do not even warrant a formal assessment. In the area of values and social implications, however, the lack of sound, effective criteria for determining which technologies to assess is decidedly evident.

Despite the abundance of offerings from bioethics on microallocation issues,* the methods for assessing values and classifying social effects have thus far received little attention in the literature. Macroallocation issues** are well described in the literature, but more apt to be ignored in an analysis. The panel that examined the development of the artificial heart, for example, focused its discussion on the microdistribution issue of which individuals would receive this scarce, expensive technology. The panel was criticized for failing to consider the artificial heart as an experimental device raising profound questions of patients' abilities to meet informed consent criteria (46). However, this concern misses an even larger social question: Is the development of technologies such as the artificial heart, which benefit only a few, a proper way to spend social resources? What then are the implications, the social costs and benefits—and how are these distributed—to society? Developing and maintaining an effective concern for the social implications of medical technology will be extremely difficult unless further work is devoted to important questions such as these.

Systematic consideration of relevant social and ethical values will not necessarily lead to conclusive answers about which policies decisionmakers should adopt. Nevertheless, choices and compromises need to be identified so that decisionmakers can see which ethical principles will be sacrificed or compromised by specific policy options. Value analysis cannot determine what the policymakers' values should be, but it can bring into focus the impact of choices on established goals and institutions (373). By describing the complex interaction of interests within the confines of the economic, legal, social, and cultural values, and technical facts prevailing at the time and anticipated in the future, value analysis can provide a more focused, conceptually clear understanding of policy problems. Value analysis can show decisionmakers where they disagree and why they disagree, as well as identify the longrun social implications of their decisions.

---

*Now the National Heart, Lung, and Blood Institute.

*Microallocation issues are concerned with single specialized economic units (e.g., individual, hospital, household).

*Macroallocation issues are related to larger or multiple economic units which make up the economy (e.g., government, business, health care).

POLICY-RELATED ACTIVITIES FOR CONSIDERING SOCIAL EFFECTS OF TECHNOLOGIES

Congress has explicitly recognized both the value of technology assessment and the importance of considering the social implications of technological change and development. This recognition is manifest, at least in part, in legislation establishing the Office of Technology Assessment (OTA), the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (National Commission), the President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research (President's Commission), and the National Center for Health Care Technology (NCHCT). The activities and efforts of these bodies are reviewed briefly below. Also reviewed are the activities of the Ethics Advisory Board in the Department of Health, Education, and Welfare (DHEW).*

Office of Technology Assessment

OTA was established in 1972 as an analytic support agency of Congress to conduct policy research on science and technology issues for congressional committees. OTA clarifies the range of policy options available on a given set of issues, and assesses the potential physical, biological, economic, legal, and social impacts that might result from adopting each option. OTA has conducted assessments in wide-ranging areas of congressional interest, including energy, international security and commerce, materials, food and renewable resources, biological applications, communication and information technologies, oceans and environment, space, transportation, innovation, and health. Although OTA reports have only rarely been directly translated into policy or legislation, they do serve to provide comprehensive background information on complex issues related to scientific and technological developments.

As exemplified by this report, OTA's health reports have primarily focused on "generic" issues in the use and assessment of medical technology. OTA has studied specific technologies (e.g., CT scanners) as illustrative issues in technology assessment and policy.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research

The National Commission was established in 1974 to develop ethical guidelines for conducting research on human subjects, and to recommend applications of these guidelines for research conducted or supported by DHEW (now DHHS). In establishing the National Commission, Congress also requested recommendations regarding the protection of human subjects in research conducted outside DHEW's purview.

The National Commission's work, which ended in 1978, was prodigious. Reports and recommended guidelines were generated covering research on human fetuses, children, prisoners, and the institutionalized mentally infirm, and on the appropriate utilization of psychosurgery. The National Commission proposed guidelines for protecting patients in DHEW-funded health care centers. It also recommended the establishment of institutional review boards in research centers as a means of ensuring that biomedical and behavioral research efforts were conducted in an ethically acceptable fashion. Many of the National Commission's proposals were recommended, revised, and adopted by DHEW, particularly those governing the protection of human subjects.

The National Commission completed two reports that have been important resource documents to later studies of human experimentation and biomedical technology. One, *The Belmont Report* (88), reviewed and clarified the ethical underpinnings of research conducted with human subjects, removing much of the conceptual confusion and semantic misunderstanding that had confounded previous attempts at rational policy for human research. Ethical analyses were developed for the distinction between research and practice, the lack of distinction between therapeutic and nontherapeutic research, the notion of risk, and the purpose and use of informed consent.

---

*Now the Department of Health and Human Services (DHHS).

The National Commission also conducted a special study of the social, ethical, and legal implications of advances in biomedical and behavioral research and medical technology (89). The study addressed the implications of computer applications to medicine, life-extending technologies, genetic screening, and reproductive engineering. The report offered no definitive recommendations and was less easily translated into policies than some of the commission's other works (349).

Ethics Advisory Board

One of the National Commission's recommendations to the Secretary of Health, Education, and Welfare resulted in the establishment of the Ethics Advisory Board in 1978. The board was given a broad mandate to review ethically questionable research protocols and research involving human subjects. Its most formidable efforts focused on in vitro fertilization and embryo transfer (the National Commission determined that under certain specified conditions, such research could be conducted in an ethically acceptable manner). The board also examined ethical questions raised by use of fetoscopy for diagnosing sickle cell anemia and hemoglobinopathies.

Because it was established in the Office of the Secretary, the Ethics Advisory Board often fielded queries from other DHEW agencies. Conducting research with human subjects raised particular problems for agencies in handling inquiries generated under the requirements of the Freedom of Information Act. The Centers for Disease Control (CDC) and the National Institutes of Health (NIH) were especially concerned about releasing specific types of research and epidemiologic data, because the data were supplied voluntarily by health care providers and institutions. CDC was afraid that its sources would dry up if the data were released. NIH wanted to avoid disclosing incomplete or preliminary findings from its ongoing clinical trials and observational studies. The Ethics Advisory Board focused attention on the need for those agencies to withhold information under the provisions of the Freedom of Information Act. Funding for the board was eliminated in 1980.

President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research

The President's Commission was established as the successor to the National Commission in 1978. Its enabling legislation directs the Secretaries of Health and Human Services and Defense, the Administrator of the Veterans Administration, and Directors of the Central Intelligence Agency, National Science Foundation, and White House Office of Science and Technology Policy to appoint representatives as liaisons with the President's Commission. Thus, the President's Commission's sphere of influence is much broader than was that of the Ethics Advisory Board.

The President's Commission is mandated to conduct studies in the broad areas of medical practice and biomedical research. The President's Commission is also to examine five specific subjects for their legal and ethical implications and their importance to public policy:

1. the requirements of informed consent for participation as human research subjects and for receiving medical treatment;
2. the procedures designed to assure the privacy of human subjects, the confidentiality of individually identifiable patient records, and appropriate access for patients to information contained in medical records;
3. the issue and advisability of developing a uniform definition of death;
4. voluntary screening, counseling, information, and education programs concerned with genetic diseases, and the fundamental equality of all human beings, born and unborn; and
5. the differences in availability of and access to health services as determined by such variables as income and residence.

In its first year, much of the President's Commission's work focused on the protection of human subjects. It examined proposed DHEW regulations (which had been based largely on the guidelines of the National Commission), specific problems inherent in social science research, and the operations of institutional review boards. The President's Commission paid particular attention to an issue originally raised by the Ethics Advisory Board: compensation for subjects injured in biomedical and behavioral experimentation. This is the only activity of the former Ethics Advisory Board carried on by the President's Commission; the President's Commission does not review questionable protocols for DHHS-funded research.

The first formal report of the President's Commission was released in July 1981 (296). A landmark examination of the medical, legal, and ethical issues surrounding the determination of death, the report is expected to serve as a model in formulating a uniform statute defining death.

In keeping with the legislative requirement that it respond to specific Presidential requests, the President's Commission has been studying issues related to genetic engineering. At the urging of religious leaders, the President's Commission was asked to address the social and ethical implications of the new technology. A draft report, focusing on questions of safety, technical capabilities, and specific issues in therapeutic and diagnostic use, was in preparation at the time of this writing.

The President's Commission has spent most of the past year looking at the distribution of health care resources, particularly as it differs among population groups. To date, the discussion has included barriers in access to care, disparities and differentials in the utilization of health services, an attempt to identify relevant ethical principles aimed at defining equity of access, the nature of special health care facilities, cost considerations, and freedom of choice for both patients and providers.

National Center for Health Care Technology

NCHCT was established in 1978 to undertake and support assessments of medical technologies, including questions of their safety, effectiveness, and cost effectiveness and their economic, ethical, legal, and social implications. Funding for NCHCT was not provided for fiscal year 1982. Congress had envisioned two primary missions for NCHCT: 1) stimulating increased scrutiny of new and existing health care technologies to ensure that the questions listed above were thoroughly explored; and 2) encouraging the dissemination of new technologies proven safe, effective, and cost effective (277).

NCHCT had no regulatory authority. Its purpose was to provide current evaluations of health care technologies to individuals and agencies with regulatory and decisionmaking responsibilities. NCHCT funded and conducted two types of assessments—"focused" and "full"—but did no actual testing of technologies.

"Focused" assessments examined the scientific and medical aspects of a technology to evaluate the technology's safety and efficacy. Such assessments were prompted by coverage questions raised by HCFA. NCHCT gathered and evaluated data, then made a recommendation to HCFA as to whether the technology should be covered by Medicare. By 1981, NCHCT had completed more than 60 focused assessments.

"Full" assessments examined the technology's safety, effectiveness, cost effectiveness, and social implications of a technology. Full assessments were integrated analyses conducted by NCHCT in cooperation with other interested Federal agencies, representatives of appropriate private sector organizations, and individuals from a broad range of relevant disciplines. When possible, participants attempted to reach agreement and provide recommendations for appropriate utilization of the technology. NCHCT identified technologies for full assessments with the assistance of an advisory council. Such assessments were conducted for coronary artery bypass surgery, cesarean delivery, and dental radiology.

CONCLUSION

Value analysis has a dual role in medical technology assessment. The first is to consider the effects on society's cultural, ethical, and political values that may result from the introduction, modification, or extension of medical technologies. Such effects cannot easily be measured and balanced, yet can profoundly affect determinations of a technology's worth. Like any form of policy analysis, technology assessment is founded on value premises. The second role of value analysis is to ensure that these value premises are made explicit and do not unduly influence the outcome of the assessment.

There is no established set of methods or techniques for conducting value analyses, nor is there a coherent, agreed upon set of principles an analysis should incorporate. Government efforts to promote value analysis probably do not require coherent sets of methods or principles. In considering the social implications of medical technology, Government promotes more comprehensive policy analysis; and in making the value premises of assessments explicit, it furthers accountability for its decisions.

Methods for synthesizing information about the health effects of medical technologies and for combining this information with information about such technologies' economic and social effects are discussed in the next chapter. Also discussed are methods for dealing with uncertainty.

5.

Synthesis, Cost-Effectiveness Analysis, and Decisionmaking

> Truth is rarely pure, and never simple.
>
> —*Oscar Wilde*

Contents

Page
Introduction 57
Synthesizing Research Results 57
  The Literature Review: Unstructured Synthesis 57
  Structured Synthesis Procedures 58
    Classification or Voting Method 58
    Meta-Analysis 58
    Other Synthesis Procedures 59
    Advantages of Structured Synthesis Procedures 60
Cost-Effectiveness Analysis 60
Group Decision Methods 62
  Delphi Technique 62
  Nominal Group Technique 62
  Consensus Development 63
  Knowledge Base Development 63
  Advantages and Disadvantages of Formal Group Process Methods 65
Decisionmaking Under Uncertainty 66
Conclusion 67

List of Tables

Table No. *Page*
1. NIH Consensus Development Meetings, September 1977 Through October 1982 64

5.
Synthesis, Cost-Effectiveness Analysis, and Decisionmaking

INTRODUCTION

This chapter is concerned with synthesizing research information and economic data in a systematic fashion so that results are useful for policy. This is not an easy task. There is an overwhelming amount of information on medical technologies (basic research, applied research, cost/outcome data), and policy analysts are faced with the problem of making sense of this information. This chapter describes and critiques methods that are intended to assist policy analysts in weighing information from diverse sources. Much of the discussion is based on material presented in appendix C and previous OTA reports (270,271).

The first section discusses methods for synthesizing research results from multiple studies which are generally concerned with safety and efficacy. Next, economic concepts are included in a presentation of the principles of cost-effectiveness analysis (CEA). This discussion is followed by a presentation of methods for soliciting group opinions and a brief discussion of quantitative decisionmaking techniques.

SYNTHESIZING RESEARCH RESULTS

Individual research studies, in themselves, do not constitute a technology assessment. An assessment must consider the evidence from a set of studies, evidence concerning social effects as well as safety and efficacy. Typically, the first step in conducting a medical technology assessment is a review and synthesis of available research evidence. Although issues pertaining to synthesis have been neglected in the health care literature (266,398), consideration of these issues is essential if the range of information available about particular medical technologies is to be utilized. Problems encountered in using traditional synthesis procedures such as the literature review are described below. Also discussed are various formal procedures to systematically synthesize research results.

The Literature Review: Unstructured Synthesis

The traditional approach to synthesizing research is the literature review. Typically, a reviewer selects a set of studies believed to be most relevant and summarizes the evidence. There are a number of problems in relying on literature reviews. First, such reviews tend to be subjective. Second, methodological problems in individual studies are often ignored, distorted, or obscured. Finally, there is the problem of timeliness: the reviews must be available prior to the dissemination of ineffective or unsafe treatments.

A number of biases may affect reviews of the research literature. The understandable enthusiasm which investigators have for any new treatment that potentially improves care can lead to errors. This problem may be somewhat diminished when reviews are prepared by groups (e.g., National Institutes of Health (NIH) panels) rather than by individual physicians or researchers.

Aspects of the problem of investigator bias are illustrated by the controversy over electronic fetal monitoring (EFM). Different perspectives apparently cause reviewers of the literature to attend to different aspects of the evidence (see app. C). The paucity of randomized clinical trials (RCTs) generally available (400) presents another important obstacle to synthesis, because well-controlled research methods are probably the best mechanism to guard against investigator bias. When such methods are employed, a reviewer can use methodological arguments to discuss disagreements between studies.

However, even when well-controlled trials are available, they may not fully answer questions about a technology. Tonsillectomy is a case in point. There exists a substantial literature on the safety and efficacy of tonsillectomies, including clinical trials and other types of research. According to Cochrane (60), none of the three different clinical trials on tonsillectomies conducted in England during the 1960's resolved the policy controversy over the appropriate use of tonsillectomy. The available RCTs, he noted, had two methodological problems: 1) the treatment was compared with no or inadequate treatment (instead of an alternate treatment); and 2) the patients' parents were not blind to the conditions of the experiments, so those whose children were on the waiting list may have exaggerated their children's symptoms.

Structured Synthesis Procedures

Because of the problems inherent in literature reviews, efforts have been made to develop more systematic procedures to integrate and interpret sets of research evidence. These range from elementary classification procedures to sophisticated statistical techniques. A description of such techniques is presented below. Additional material is presented in appendix C.

Classification or Voting Method

A simple structured synthesis technique involves organizing a body of literature according to a prespecified set of criteria and is actually a classification procedure (226) Sometimes called the "voting method," this synthesis technique involves selecting a particular sample of evaluative studies of a technology, coding some aspect of the design and/or conceptual framework, classifying observed outcomes as to whether they are favorable, neutral, or unfavorable (i.e., "taking a vote"), and then constructing tables of research findings. This method is frequently used to demonstrate the differences obtained by various methodological approaches used to assess the same technology.

The value of the voting method lies in the precise identification of the type of studies to be sampled, and the coding of substantive and methodological aspects of the studies according to clearly defined procedures. More widespread use of this classification technique could probably aid in determining the specific patient populations and/or conditions that can be effectively treated by a medical technology (i.e., to establish external validity). This technique helps to avoid the problems of more traditional literature reviews, noted above, which selectively describe research evidence and which neglect consideration of methodological strengths and weaknesses.

Krol (216), however, cites three problems with the method: sample size, effect size, and Simpson's paradox. The first problem is that large studies are likely to produce statistically more significant results than studies with small numbers of subjects. Thus, for example, a study finding of "no results" or "no difference" between treatment and control conditions may be correlated with small sample size. The second problem is that the voting method does not take effect size into account, i.e., small, marginal effects are not distinguished from large effects. Simpson's paradox is a more subtle statistical point. It is possible, under certain conditions, to reach conclusions by aggregating data from all studies that are different from the conclusions reached by counting each study separately. A fourth problem with the voting method is that it may oversimplify the results of studies and cause reviewers to overlook subtle interactions among variables (226).

Meta-Analysis

A rigorous statistical approach to research synthesis is a sophisticated quantitative synthesis technique called meta-analysis (166,165). This technique uses the actual results of studies and permits the determination across a set of studies of the magnitude of treatment impact. Meta-analyses are useful in assessing treatments for which a large number of studies are available and findings across studies seem to have great variability.

As used by Glass (165), such analyses require that both comparison and treatment groups be available and that the original research reports contain appropriate statistical information such as the group mean and standard deviation. Effect sizes (ES) are calculated by determining the difference between the mean of the treatment group (T) and the mean of the comparison group (C), divided by the standard deviation of the comparison group (SD). Thus,

$$ES = \frac{\overline{T} - \overline{C}}{SD}$$

This procedure converts the average effect of each outcome measure into a common scale (i.e., standard deviations) that can be compared to results of other studies. If a treatment has no effect, then the effect size would be zero; if the treatment is effective (i.e., better than the current alternative), then the effect size is positive; and if the treatment is inefficacious, the effect score is negative. If some assumptions are made about the skewness of experimental and control group scores within each study and the distribution of effect sizes across a large number of studies (i.e., that they are normally distributed), effect sizes can be converted into percentile ranks and inferences can be made about the overall effects of a medical technology.

One recent example of the use of meta-analysis is Smith, Glass, and Miller's (353,354) review of the outcome literature on psychotherapy. These investigators searched the published literature and included within their analysis all available control group studies of the effectiveness of any form of psychotherapy.* For each study, the investigators coded an extensive number of variables, including methodological criteria such as the nature of patient assignment to condition (e.g., random v. matching), experimental mortality, and other threats to internal validity. Effect size scores were calculated for each principal outcome. A code was also developed for validity of the outcome measures.

Smith, Glass, and Miller's (354) findings indicated that, on average, the difference between scores of the groups receiving psychotherapy and the control groups was 0.85 standard deviation units. Assuming the normal distribution of effect size scores, this average standard score can be translated to indicate that a typical person who receives psychotherapy is better off than 80 percent of the persons who do not. The investigators also performed a number of analyses to determine whether the methodology of the outcomes study affected results and whether different therapies (or other factors) were differentially efficacious.

The work of Smith and colleagues (353) has been criticized, because it "lumps together" a large number of what some consider incomparable treatments and outcomes (e.g., 137). It should be noted, however, that the strength of the effect size technique is that it provides a common metric that permits analysis of these differences (methodological and substantive). Smith, et al.'s (354), classification variables for each study were fairly comprehensive and yielded a systematic comparison of studies on the basis of their conceptual and methodological designs. What is problematic, however, is that the findings are heavily dependent on a number of decisions that are not always made explicit. These include criteria for selecting literature and criteria for selecting variables. It is not possible to ascertain biases resulting from the investigators' sampling decision and whether only certain types of studies, therapies, or variables are assessed using control group designs (273).

Other Synthesis Procedures

A number of other methods exist for statistically combining the results of independent studies (see 69,292,324). The effect size method described above actually incorporates several procedures. The most important of these is the comparison of treatments to detect interactions between the characteristics of a study and outcome (i.e., external validity). As noted in the discussion of the classification method, some of these procedures can be employed when effect scores are not computed.

Additional statistical methods can be used to combine probability values from various studies and to adjust outcome scores according to the relevance of the data. Rosenthal (324) describes a number of procedures for combining probabilities. These range from adding observed probability levels across different studies to adding weighted standardized scores. They also include the testing of mean probability values. Use of such procedures indicates whether significant effects are obtained across a set of studies. The problem in using probability values is that the number of subjects per study influences the statistical power to detect whether significant overall differences are present.

---

*Drug studies were analyzed separately, and studies that did not involve the use of professional therapists (operationally defined as psychologists, psychiatrists, and social workers) were eliminated from the analysis.

An interesting method for synthesizing the results of experiments done with human and animal species according to the relevance of the data has been proposed by Du Mouchel and Harris (131). This method involves the sophisticated application of a statistical theorem (Bayes' theorem) to provide a quantitative prediction of data relevance. Du Mouchel and Harris use the procedure to provide estimated carcinogenic risk of various substances derived from the results of a series of epidemiological studies.

Advantages of Structured Synthesis Procedures

A number of advantages result from the use of formal procedures for data synthesis (292). One benefit is that formal syntheses help to identify contradictions in the literature by systematically organizing studies according to specified classification factors. Thus, differential outcomes can be segregated according to treatment chacteristics and/or methodological approaches.

The use of effect size scores offers a second benefit. Such scores provide insight as to the worth of a treatment and provide a benchmark for later research. Thus, for example, an analysis of 23 controlled studies of patient education programs by Posavac (294) found a 0.75 average effect size. According to Posavac, this should provide a standard against which new patient education programs can be assessed. A finding that the effect size of a new program is only 0.20 (providing that similar dependent measures are employed), would probably indicate that the program was not particularly effective, at least for the problem or population for whom it had been designed.

Another advantage of quantitative synthesis methods is that they serve to control for some statistical conclusion validity problems (e.g., power) that some commentators have reported as severe in the medical literature (e.g., 141,172,337). The widespread use of meta-analysis and other quantitative approaches to research synthesis would likely improve statistical reporting practices by calling attention to investigators' use of data. Furthermore, most quantitative analyses of multiple studies would compensate for errors in analyses such as the use of multiple-independent inferential tests without appropriate error rate control or incorrect inferences because of a lack of power. Although errors in data collection would not be corrected by quantitative synthesis methods, the systematic analysis of multiple studies should render the effect of such errors less consequential. The attention to systematic considerations of the "weight" of evidence across research studies should have a generally salutary effect.

These synthesis procedures seem most appropriate for evaluating more mature medical technologies about which there has accumulated a considerable body of research. Often however, they may be applicable to less developed technologies. In some cases, where only meager evidence is available from a small set of studies, the effectiveness of a medical technology may be suggested by a review of specific components from some other portion of the literature.

COST-EFFECTIVENESS ANALYSIS

CEA can be thought of as an aid to synthesis of both the health effects and the economic effects of a medical technology. CEA was itself the topic of a recent OTA assessment (270,271,272, 273,274). OTA's findings from that study are summarized here.

The value of CEA lies more in the process of performing the analysis than in any numerical results. There are a number of reasons for this, among the most important of which are CEA's inability to adequately address ethical issues and the uncertainty of many of the key variables required for the analysis. Addressing uncertainty is a topic addressed later in this chapter.

A second major finding from the OTA study was that there is no one "correct" way to do an analysis. Not only does each analysis differ in terms of which benefits and which costs must be considered, but each analysis differs in terms of how the benefits and costs are valued. A driving force behind these variations are the social/ethical concerns mentioned earlier.

OTA suggested that the most appropriate approach to any assessment is to perform it in an open forum so that assumptions and underlying values can be challenged; to identify, measure, and, to the extent possible, value all relevant benefits/effects and costs; and to present the results of the analysis as an "array" of benefits/effects and costs rather than forcing them into some aggregate single measure. By arraying effects in a systematic fashion, one can place the appropriate relative emphasis on given effects whether they are quantifiable or not. This technique is designed to make more explicit the health, economic, and social consequences of any decision.

In suggesting the "array" method, OTA recognized that CEA is a decision-assisting technique, rather than a decisionmaking one. In some instances, however, a cost per aggregated effect may be possible, appropriate, and quite acceptable. A case in point is OTA's own cost-effectiveness study on pneumococcal vaccine, which calculated a ratio of $1,000 per quality-adjusted year of life saved for the elderly (282). In that case, the study was performed under public scrutiny, and the analysis and assumptions were subjected to extensive outside review.

Finally, although OTA concluded that there was no single "correct" methodology for conducting CEA, it did find general agreement on 10 principles of analysis. Those principles, including a short explanation of each, are reproduced below.

1. *Define Problem.*—The problem should be clearly and explicitly defined and the relationship to health outcome or status should be stated.
2. *State Objectives.*—The objectives of the technology being assessed should be explicitly stated, and the analysis should address the degree to which the objectives are (expected to be) met.
3. *Identify Alternatives.*—Alternative means (technologies) to accomplish the objectives should be identified and subjected to analysis. When slightly different outcomes are involved, the effect this difference will have on the analysis should be examined.
4. *Analyze Benefits/Effects.*—All foreseeable benefits/effects (positive and negative outcomes) should be identified, and when possible, should be measured. Also, when possible, and if agreement can be reached, it may be helpful to value all benefits in common terms in order to make comparisons easier.
5. *Analyze Costs.*—All expected costs should be identified, and when possible, should be measured and valued in dollars.
6. *Differentiate Perspective of Analysis.*—When private or program benefits and costs differ from social benefits and costs (and if a private or program perspective is appropriate for the analysis), the differences should be identified.
7. *Perform Discounting.*—All future costs and benefits should be discounted to their present value.
8. *Analyze Uncertainties.*—Sensitivity analysis should be conducted. Key variables should be analyzed to determine the importance of their uncertainty to the results of the analysis. A range of possible values for each variable should be examined for effects on results.
9. *Address Ethical Issues.*—Ethical issues should be identified, discussed, and placed in appropriate perspective relative to the rest of the analysis and the objectives of the technology.
10. *Discuss Results.*—The results of the analysis should be discussed in terms of validity, sensitivity to changes in assumptions, and implications for policy or decisionmaking.

GROUP DECISION METHODS

Although the application of formal quantitative procedures for the integration of data from individual studies and the use of quantitative decision-assisting techniques should improve the process of technology assessment, such procedures cannot, by themselves, resolve policy controversies. These procedures cannot go beyond the data available on a particular problem, nor can they substitute for informed judgment or include societal values (see ch. 4).

A frequently used method for soliciting group opinions is the unstructured conference at which a given topic (or topics) is discussed. There are many conference formats, ranging from formal to informal presentations with or without questions and answers. Sometimes prepared critiques or presentations are employed; other times a debate format is used.

Another informal group opinion technique is the advisory panel approach such as that used by many Government agencies (including OTA). Sometimes, the panel is required to endorse a given study. In such cases, studies are often modified until most members are satisfied with the major points. Often a single member or small number of members of the panel can exercise de facto veto authority over certain elements of the study; other times a minority is included in the final report.

Four formal methods for resolving conflicts across research studies and for developing assessments of particular technologies are described below: 1) the Delphi technique; 2) the nominal group technique; 3) a new group opinion process, referred to as consensus development, sponsored by NIH; and 4) a computerized knowledge base being developed by the National Library of Medicine (NLM). The Delphi and nominal group techniques are based on behavioral science principles (163), the goal being to aid groups composed of individuals with different information and perspectives to develop group judgments that best take account of the positions of the individual members. The discussion below illustrates both the potential and limitations of these methods in synthesizing technology assessment information.

Delphi Technique

Delphi is probably the oldest structured model for involving groups in decisionmaking processes and has been used widely in health care (78). The Delphi technique uses a series of questionnaires (or individual interviews), each followed by anonymous feedback summarizing all the participants' responses. Although Delphi was originally developed by the Rand Corp. to synthesize expert opinions on national defense problems, it has since been extended to medical problems (232,246,250, 318,336).

A unique feature of the Delphi technique is that persons selected to participate in the process generally have no direct contact with one another. Instead, participants are provided with a summary of the questionnaire responses, usually by mail. Personality or status variables thus have little chance to influence participants' opinions, as they might in face-to-face meetings. By using anonymous feedback, each participant has an equal chance of influencing other participants (41). The technique also provides a framework within which to approach a problem in a focused manner. Finally, and perhaps most importantly, the Delphi technique provides a limited time frame in which to achieve consensus (41,135). There are a fixed number of iterations, usually three, in the questionnaire-feedback process. (For a discussion of a modified Delphi technique, see ref. 293.)

Nominal Group Technique

Although there are several variations of this technique, typically, all participants are seated at a common table and asked to write their views on each of a number of issues posed by the leader of the meeting. Delbecq and colleagues (82) call this the "nominal" group technique, because the individuals at the table (at the outset) are a group in name only. Each person's view is recorded on a separate card, and talking is prohibited. The (silent) presence of others while writing the cards is supposed to stimulate participants to perform better. The cards are collected, and their contents (but not the authors' names) are listed for all to see. Subsequent discussions focus on the ideas that have been proposed, not on who did the proposing.

Consensus Development

In response to congressional pressure to assist in the transfer of technology, NIH initiated its consensus development program in 1977. Its goal is to bring together various concerned parties—physicians, consumers, bioethicists, etc.—to seek consensus on the safety, efficacy, and appropriate conditions for use of various medical technologies. Judgments about the technology under consideration are intended to be based on the available scientific evidence. At conferences jointly sponsored with the National Center for Health Care Technology (NCHCT), information about the technology's social, ethical, economic, and legal impacts was included, as well.* The consensus development process is designed to produce a written document, called a "consensus statement," that can be accepted by clinicians, researchers, and the public. The statement is supposed to identify both what is known and not known about the technology (287). (A list of NIH consensus development meetings from September 1977 through October 1982 appears in table 1.)

The consensus development conferences are coordinated by NIH's Office for Medical Applications of Research (OMAR). The topics are selected by the relevant institutes, but OMAR helps to make the final decision in cooperation with the bureaus, institutes, and divisions of NIH about the suitability of the topic, panel composition, and the proposed format for a consensus exercise. Adversary groups and task forces have been almost entirely abandoned. Moreover, the questions that have been posed to the conferences have been addressed strictly to those issues on which there is enough factual evidence to reach agreement. Unlike the techniques discussed above, consensus conferences have no particular theoretical bases for their format.

A panel of experts is selected by NIH to hear presentations by the leading medical researchers addressing a prespecified set of questions about the technology. These presentations, usually summarizing the latest research findings, are made over a 2-day period during which both panelists and audience members discuss the research findings. On the evening of the second day, the panel is requested to draft a statement responding to the questions. Usually, the panel deliberates through much of the night, often writing four or more drafts of its consensus statement. In some rare cases, minority reports are developed to indicate disagreement with the majority recommendations. The next morning, the consensus statement is read to the audience for their comments and criticisms. The conference concludes with a press conference. After the panel disperses, it sets about the final task of revising the statement. The statements from consensus development conferences are widely disseminated to thousands of organizations and individuals by NIH and by publication arrangements with leading medical journals such as the *New England Journal of Medicine.*

Knowledge Base Development

In response to the often overwhelming number of articles and other information concerning a particular topic, NLM's Lister Hill National Center for Biomedical Communications developed a unique system for soliciting expert opinion. NLM's system, termed the "Hepatitis Knowledge Base," is intended to function as a continually updated, computerized body of knowledge concerning viral hepatitis (129). Although this system's topic is a disease rather than a technology, its approach to soliciting and maintaining expert opinion could also be applied to the latter.

Information for the Hepatitis Knowledge Base was originally assembled by a small body of experts who reviewed and combined the critical information from an identified set of 40 important review articles on the subject. Any inconsistency in the information was addressed by these experts; they either reached a consensus or noted that there was an unresolvable conflict.

The resulting base of "knowledge" was computerized and made immediately accessible to each expert within this group. Subsequent information which appears in the literature is reviewed

---

*NCHCT, which was established in 1978, was not funded in 1982.

Table 1.—NIH Consensus Development Meetings, September 1977 Through October 1982

Sponsors	Title	Dates held
NCI	Breast Cancer Screening	Sept. 14-16, 1977
NCI	Educational Needs of Physicians and the Public Regarding Asbestos Exposure	May 22, 1978
NIDR	Dental Implants Benefit and Risk	June 13-14, 1978
NCI	Mass Screening for Colo-Rectal Cancer	June 26-28, 1978
NIA	Treatable Brain Diseases in the Elderly	July 10-11, 1978
NINCDS	Indications of Tonsillectomy and Adenoidectomy: Phase I	July 20, 1978
NIAID	Availability of Insect Sting Kits to Nonphysicians	Sept. 14, 1978
NCI	Mass Screening for Lung Cancer	Sept. 18-20, 1978
NIGMS	Supportive Therapy in Burn Care	Nov. 10-11, 1978
NIAMDD	Surgical Treatment of Morbid Obesity	Dec. 4-5, 1978
Interagency Committee on New Therapies for Pain and Discomfort (Organizer)	Pain, Discomfort, and Humanitarian Care	Feb. 16, 1979
NICHD	Antenatal Diagnosis	Mar. 5-7, 1979
NHLBI	Transfusion Therapy in Pregnant Sickle Cell Disease Patients	Apr. 23-24, 1979
NHLBI	Improving Clinical and Consumer Use of Blood Pressure Measuring Devices	Apr. 26-27, 1979
NCI	The Treatment of Primary Breast Cancer: Management of Local Disease	June 5, 1979
NCI	Steroid Receptors in Breast Cancer	June 27-29, 1979
NEI	Intraocular Lens Implantation[a]	Sept. 10-11, 1979
NIA	Estrogen Use and Post-Menopausal Women	Sept. 13-14, 1979
NIAID	Amantadine: Does It Have a Role in the Prevention and Treatment of Influenza?	Oct. 15-16, 1979
DRS	The Use of Microprocessor-Based "Intelligent" Machines in Patient Care	Oct. 17-19, 1979
NIDR	Removal of Third Molars	Nov. 28-30, 1979
NHLBI	Thrombolytic Therapy in Thrombosis	Apr. 10-11, 1980
NINCDS	Febrile Seizures	May 19-21, 1980
NCI	Adjuvant Chemotherapy of Breast Cancer	July 14-16, 1980
NCI, NIA, NICHD[a]	Cervical Cancer Screening: The Pap Smear	July 23-25, 1980
NIAMDD[a]	Endoscopy in Upper GI Bleeding	Aug. 20-22, 1980
NICHD[a]	Childbirth by Cesarean Delivery	Sept. 22-23, 1980
NCI	CEA and Immunodiagnoses	Sept. 29 - Oct. 1, 1980
NHLBI[a]	Coronary Artery Bypass Surgery: Scientific and Clinical Aspects	Dec. 3-5, 1980
NINCDS	The Diagnosis and Treatment of Reye's Syndrome	Mar. 2-4, 1981
NINCDS	CT Scanning of the Brain	Nov. 4-6, 1981
NIAID	Defined Diets and Childhood Hyperactivity	Jan. 13-15, 1982
NIAID	Total Hip Joint Replacement	Mar. 1-3, 1982
NIAID	Immunology—The Bee Sting	Oct. 6-8, 1982[b]

[a] Jointly sponsored with the National Center for Health Care Technology.
[b] Planned date.

SOURCE: National Institutes of Health, Office for Medical Applications of Research, 1982.

individually by the panel, consensus is sought, and the knowledge base is updated. The panel can caucus through a technique known as "computer conferencing" without being together geographically or temporally. As time permits, each member can access the terminal to enter his or her comments and to review the comments of others.

To date, the Hepatitis Knowledge Base system for reaching consensus on a timely, complex biomedical topic is a research effort only. It has not been evaluated for either clinical utility or cost effectiveness.

An important outgrowth of the Hepatitis Knowledge Base research effort is the development of the Knowledge Base Research Program. This experimental program, which currently includes viral hepatitis, peptic ulcer, and human genetics, may contribute to more effective access to and use of available biomedical information in solving the daily problems of diagnosis, prognosis, and treatment of illness. The program is exploring ways by which medical computer scientists and medical subject matter experts can select and organize relevant and accurate information from the biomedical literature.

Advantages and Disadvantages of Formal Group Process Methods

The methods discussed above are the better known and developed methods designed to facilitate group decision processes and make them more efficient. These methods are also intended to produce more reliable and valid information than an unstructured conference. Evidence of effectiveness, however, is somewhat contradictory for the Delphi and nominal group techniques, and sparse for the newer methods such as NIH's consensus development conferences and NLM's knowledge base.

The relative strengths and weaknesses of both the Delphi and nominal group process methods have been summarized by Delbecq, et al. (82). They believe that these methods are superior to simple, unstructured group interaction providing for a higher level of independent thinking both in terms of quantity and quality, especially with respect to specificity. Delphi seems to be particularly relevant for generating predictive information (78) when data are poor and for resolving highly controversial issues likely to be distorted when participants interact with one another.

Nevertheless, in comparison to formal decision analysis, the Delphi technique has been criticized as being little better than the "seat-of-the-pants" methods currently employed by policymakers and as being a method which bases "knowledge" on an informal set of opinions (332). Others (10) maintain that it is subject to the same total error as most predictions. The process is also time- and group-dependent, because the results are based on information available to a specific group of experts at a specific point in time. As a consequence, the process should be repeated as data change with time. It also appears less well suited as a process for resolving minimally controversial issues (318) or for synthesizing the state of the art in a given field (163).

A recent study compared Delphi with the nominal group process technique (368). Physicians were randomly assigned to one of three Delphi or nominal group technique panels to develop procedures for handling four hypothetical emergency medical services cases. In order to determine the reliability of the decisions, panelists were contacted individually 6 months later and asked to cast an anonymous vote on the procedures originally discussed. The degree of consensus achieved was the same for both techniques. The most striking finding, however, concerned the reliability of decisions over time. There were "very extensive" changes in the nominal group technique vote 6 months later, suggesting that it is "a less than reliable technique for reaching a consensus." Although the physicians reported that they liked the nominal group technique much more than Delphi, group norms and pressures were developed using the nominal group technique that produced unstable or false consensus.

Virtually no critiques of NIH consensus conferences have been published to date. However, a major study is currently being funded by NIH to evaluate its process. A recent Institute of Medicine (IOM) publication notes that consensus conference results were reported to be particularly useful in health planning, quality assurance, and setting reimbursement (197). Although no current evidence is available regarding the usefulness and impact of the conference results, NIH is planning to fund an ambitious impact study. Wagner suggested that the success of any consensus format is dependent on the following considerations (197):

1. the composition of the evaluation panel, especially participation by epidemiologists and biostatisticians;
2. the amount, quality, and comprehensiveness of available data;
3. the duration of the process (sufficient time for participants to synthesize information is critical; several meetings held during a longer

period are preferable to an intensive, one-time meeting); and

4. the resources for support.

Although highly structured, the NIH consensus format, is not designed, as the Delphi and nominal group techniques are, to limit group dynamic problems such as potential dominance by selected individuals or groups within the panel. From a methodological perspective, two aspects of the NIH consensus development process are of concern: its sensitivity to the limitations of the research evidence and the extent to which it considers a comprehensive and systematic review of the research literature. However, NIH is aware that its search for consensus resolution must be confined within the limits of the expertise and evidence assembled. In order to broaden the expertise and evidence to the maximum extent practicable, NIH strives for diverse, open, and balanced representation and participation at its consensus conferences. It also performs an exhaustive review of the research literature to compile relevant background evidence. A more complete discussion of NIH consensus meetings is presented in appendix C.

Relatively less is known about the value of the knowledge base approach. Nevertheless, the concept seems extremely interesting and potentially quite helpful to researchers and clinicians alike. One major concern is the cost of such a system.

DECISIONMAKING UNDER UNCERTAINTY

Elements of uncertainty abound throughout the process of assessment. Over the years, various techniques have been developed to assist in making rational decisions under such conditions. The discussion here is intended to introduce the reader to the more useful techniques for evaluating uncertainty within the field of medical technology assessment. It is not an exhaustive description of such methods, nor is it intended to provide the reader with a complete understanding of any one technique.

It is important to distinguish between two types of uncertainty: that which reflects the presence of random events and that which reflects a basic lack of knowledge. Random events occur according to a known probability distribution. For example, the flipping of a coin will result in heads roughly half of the time and tails the other half. Thus, although the result of a single flip of a coin remains uncertain, the probability distribution of heads and tails is known. By contrast, when uncertainty reflects a lack of knowledge, one not only does not know the probabilities of various outcomes' occurrence, one may not even know which types of outcomes can occur. Thus, administering a brand-new chemotherapeutic agent to a terminal cancer patient may have any of several therapeutic and toxic effects; conceivably, the latter might include side effects never previously recognized in cancer chemotherapy.

Uncertainty due to random events can be adequately handled by a variety of techniques. A common approach is statistical confidence limits. Another common method is decision analysis.* Possible courses of action (outcomes, etc.) are diagramed on a "decision tree." Branches in the diagram are associated with known or imputed probabilities and "payoffs." As a result, an outcome value (positive and negative) is associated with each pathway. Thus, analysts can trace plausible paths to determine the probability and expected value of each final outcome. (For a discussion of decision analysis in clinical decisionmaking, see ref. 386.)

In addition to decision analysis, a variety of computer simulation techniques allow analysts to model real-world phenomena and estimate their consequences over hypothetical periods of time. By manipulating all such models until outcomes mirror empirical findings, analysts may be able to acquire valuable insight into real-world processes. The potential usefulness of modeling techniques is great, but analysts and policymakers should always retain an awareness of the influence of underlying assumptions. Technical sophistication can mask tenuous assumptions, particularly for those individuals who lack familiarity with the analytical approaches.

---

*Decision analysis is also helpful for uncertainty due to lack of knowledge.

One general approach to handling uncertainty is called sensitivity analysis, which is a conceptually simple but powerful tool with which to address both random and nonrandom events. Actually a series of techniques, sensitivity analysis can test whether variations in assumptions affect the qualitative conclusions of an analysis. It is particularly useful in CEA. For instance, if an analyst assumes a discount rate of 2 percent and concludes that the program in question is desirable (i.e., its benefits exceed its costs), he or she can try discount rates of 0 to 4 percent to determine whether the program's basic desirability is a function of, or is sensitive to, the discount rate. Thus, sensitivity analysis can shed light on the importance of certain assumptions, especially as to whether an analysis is meaningful despite the presence of uncertainty.

Sensitivity analysis can produce four important results. First, it can demonstrate that a conclusion of a study substantially depends on a particular assumption, thereby suggesting that the overall analysis cannot be viewed as "definitive." Second, it can demonstrate that a conclusion is "robust" with respect to a particular assumption (i.e., that violation of the assumption does not significantly affect the conclusion) and, hence, that the tenuousness of the assumption is not a source of concern. Third, it can establish a minimum or maximum value which a variable must have for a program to appear worthwhile. Finally, sensitivity analysis can identify issues (uncertainties) deserving of research attention.

CONCLUSION

The techniques for synthesizing research, CEA, soliciting group opinions, and decisionmaking all have the common goal of making sense out of a body of information.

In technology assessment, the objective of synthesizing the information gained from multiple research studies is to understand the cause and effect relationships from the use of a technology. As chapter 3 explained, research studies, no matter what type of design, never provide perfect information and often provide only insights into relationships. Most of the synthesis techniques described in this chapter were developed to provide analyst/researchers with systematic means to make those relationships clear. These techniques are primarily concerned with safety, efficacy, and effectiveness.

CEA can be regarded as adding at least one more dimension to the synthesizing techniques described above. Traditionally, CEA has been used to balance the safety/efficacy/effectiveness of a technology with its economic effects. However, a recent OTA report, *The Implications of Cost-Effectiveness Analysis of Medical Technology* (270), described how social values can also be incorporated into the analytical framework. The 10 principles of analysis, described by OTA, are intended to make CEA more policy-relevant than most applications of the technique are. But, as OTA concluded, CEA is a *decision-assisting* rather than a *decisionmaking* tool.

A systematic approach to decisionmaking is to solicit group opinions. This approach can be used to synthesize diverse individual and group values with clinical research findings and economic effects. Several of the group process techniques described in this chapter were developed and refined by the social science community to maximize the benefits and minimize the liabilities of group dynamics. Evidence indicates that these formal techniques produce more and better information than unstructured conference-type sessions. To date, however, there have been few broad applications of these techniques.

One purpose of assessing medical technologies, is to produce information needed by policymakers. Policies that affect the development, diffusion, and use of medical technologies—in particular, drugs, medical devices, and medical and surgical procedures—are described in the next chapter.

6.

Factors Affecting the Development, Diffusion, and Use of Medical Technologies

> Technology is not an historical institution like science and capitalism, but is rather the essence of man in an active and uncritical state.
>
> —*H. T. Wilson*

Contents

Page
Introduction 71
Innovation Process for Medical Technologies 71
  Sources of Medical Technologies 71
    Drugs 71
    Medical Devices 72
    Medical and Surgical Procedures 73
  Influences on the Diffusion of Innovations 73
Programs and Policies Related to the Development, Diffusion, and Use of Medical Technologies 74
  Research Activities 74
    Federal Research Activities 75
    Private Sector Research Activities 77
  Regulatory Responsibilities 78
    Federal Regulation of Drugs 78
    Federal Regulation of Medical Devices 79
    Other Regulatory Activities 80
  Reimbursement Policies 81
    Variations of Reimbursement Mechanisms 81
    Influence of Reimbursement on the Development, Diffusion, and Use of Medical Technologies 82
  Coordination Efforts and Dissemination of Information 84
    Federal Activities 84
    Other Processes 85
Conclusion 86

List of Tables

Table No. *Page*
2. Number and Amount of NIH Support for Clinical Trials Active in Fiscal Year 1979 by Institute 75
3. Amount of NIH Support for Clinical Trials Active in Fiscal Year 1979 by Institute for Type of Support 76

6.

Factors Affecting the Development, Diffusion, and Use of Medical Technologies

INTRODUCTION

Information on medical technologies is needed by various private and public parties at every stage in the development and diffusion of medical technologies. All markets need good information to function effectively. However, the market for medical technologies has some unique characteristics that affect the type of information needed and the timing of that information.

Because of the many social values related to medical care, there is a degree of public responsibility for ensuring that medical technologies are safe, efficacious, and, in some cases, cost effective. Various Federal agencies provide research funds, regulate market entry, and decide which medical technologies will be reimbursable for Federal beneficiaries. Technology assessment information is often needed in order to make responsible decisions that protect the public safety and the public purse, yet do not unduly impede the innovation process.

This chapter is intended to set the stage for a critique of the current "system" for assessing medical technologies. The first section summarizes selected aspects of the innovation and diffusion process for drugs, devices, and surgical and medical procedures. The second section explores research, regulatory, and reimbursement policies that affect the development, diffusion, and use of medical technologies.

Additional material concerning the drug and device industries and the Federal policies that affect the innovation process is presented in appendix D. Appendix E contains five case studies of medical technologies which are intended to illustrate some of the points made in this chapter, especially regarding the way in which technologies are developed and adopted and the effects which Federal policies have on the process.

INNOVATION PROCESS FOR MEDICAL TECHNOLOGIES

In some respects, the innovation process for medical technologies parallels that for other technologies. Although there are many variations, a basic model of the process can be outlined. An innovation is conceptualized by recognizing both technical feasibility and potential demand. If a decision is made to pursue the innovative idea, problem-solving activity follows, drawing from available information and from further research and development (R&D) activities. If a solution to the problem is found, it may be the one originally sought, or a solution to a modification of the original problem.

Sources of Medical Technologies

Drugs

The U.S. pharmaceutical industry is composed of about 600 firms. These include a small number of large firms which produce most of the innovative drugs and a much larger number of relatively small firms which produce and market mainly generic and some patented drugs. The industry is characterized by high and rising development costs for new products, and there has been a strong shift toward greater concentration of new products in the largest firms.

Since the late 1950's, the number of new chemical entities and number of firms producing a new chemical entity has declined (174). Innovative outputs have been concentrated in the 20 largest of the 600 drug firms, and most of this concentration is among the top four to eight innovators. From 1957-61 to 1967-71, the four largest firms in the industry nearly doubled their share of innovative output. During the same period, however, their share of total prescription drug sales remained fairly constant.

These observations suggest that most of the large drug firms are dependent on a few drugs for much of their income. Apparently, after the patents for new products expire, generics erode some of the market captured by the innovating drug firm, and the large innovative firms regain their share of total sales through the introduction of new drugs.

The Food and Drug Administration's (FDA's) regulatory responsibilities regarding new drugs are discussed in the next major section of this chapter.

Medical Devices

The U.S. medical devices industry has experienced substantial growth since World War II. Industry sales in 1977 were $8.1 billion—five times the amount in 1958 (corrected for inflation). Growth has been predominantly in the number of firms rather in their size. The U.S. medical devices industry is composed of several thousand firms—many specialized small firms which together have a small share of the market and a few large firms with a high market share. There are high entry and exit rates in the industry, mostly among small firms (8).

Dominance by large companies suggests the presence of economies of scale, while the persistence of many small companies suggests that economies of scale do not apply to specialized areas. Possibly, however, the large firms really represent the industry; i.e., rather than representing the differentiation of the industry into small and large functions, the large number of small firms may represent a high-birth, high-mortality, and high-turnover sector of the industry (122). Arthur Young & Co.'s (8) survey of the industry, for example, did not differentiate between bankruptcy and acquisition in its observation of high-turnover rates for small firms. However d'Arbeloff (79) comments that high-turnover rates may reflect a high-risk, high-profit atmosphere for small firms.

In general, small firms fill a special niche in the market, and their growth into larger firms is hindered by conditions such as advertising requirements, links with distribution channels, and the need for new capital expenditures (355). Thus, the industrial pattern is that of limited internal growth, with acquisition or establishment of additional companies being the primary method of expansion. Small plants are opened to manufacture new products following invention and development, while large plants are opened by large companies to take advantage of lower operating costs. These large companies tend to be extremely diversified as a whole, yet there is little product diversification within their medical devices plants (8).

Recently, the distribution of medical devices has shifted from small regional and local suppliers to major national dealers. National dealers are often subsidiaries of large manufacturers or are acquirers of small manufacturing firms. The advantage of larger firms is that they are better positioned to provide special buyer education through their larger, better trained staff (355). The inability of potential manufacturers to gain access to these networks is an additional barrier to growth of the small firms entering the medical devices field and probably accentuates their acquisition by larger manufacturers.

The U.S. medical devices industry is somewhat insulated from price competition by the high level of third-party reimbursement, and price competition is not as significant a force in mitigating price increases as it is in other industries. Nevertheless, there is a high degree of product differentiation, and the industry appears to be competitive at various levels even though the market for the most part is price insensitive (8).

FDA's regulatory responsibilities regarding medical devices are discussed in the next major section of this chapter.

Medical and Surgical Procedures

The invention, development, and diffusion of medical and surgical procedures can to some degree be described by the model of the innovation process developed for products and their manufacturing processes. For the most part, however, medical and surgical procedures are developed within the practice of medicine. Furthermore, the initial diffusion of medical and surgical procedures is relatively uninhibited by Federal regulation.

New procedures usually involve some drug and/or device. However, a focus on procedures separate from the technologies which are used in them is necessary, because physicians, as users, are both generators (technology-push) and purchasers (demand-pull) of these innovations. Thus, it is important to understand how physicians perform these dual roles. But there are no standard determinants of when or how procedures become medically acceptable (197) and few criteria for when they become obsolete.

Influences on the Diffusion of Innovations

The medical literature on communication about and adoption of innovations is weighted toward studies of single diagnostic or therapeutic medical technologies. There is a large literature on how physicians learn about and adopt new drugs and a growing literature on specific devices or techniques, but very little literature on the communication about or the adoption of complex medical procedures that may not involve drugs or hardware (e.g., psychotherapy). In practice, however, the crucial distinction is between communication which informs the physician about novel technologies and that which influences physicians to act (405). Even though the most important source of new knowledge about improvements in medical technologies is the professional literature, physicians cite professional colleagues more often as sources they turn to when deciding to use a new procedure for the first time (145,233, 234).

The importance of informal communication both in the process of scientific discovery and in the diffusion of technological innovation seems to be a feature in all fields of technological discovery and diffusion (213). Moreover, it may be that there is a prestige hierarchy, where those at the top are "trend setters" (49). If this is so, widespread adoption of an innovation could be enhanced by convincing influential organizations to adopt it first, then letting prestige-seeking organizations imitate them.

Physicians of greater prestige do tend to hear about innovations sooner than others (62), and they are also mentioned by their fellow professionals as influential sources of information on the medical practice of others. However, the adoption process when the adopting unit is an organization (e.g., hospital) is substantially different from the process when the adopting unit is an individual (e.g., physician in solo practice) (178,405), and these processes differ by the level of complexity of the organization. Outside forces such as third-party reimbursement or regulatory practices may also affect how quickly the individuals in the medical community learn about or adopt a technology.

These theoretical and empirical findings point to a kind of general scenario. Medical and surgical procedures usually begin as user- (e.g., physician) generated innovations. In medicine, an innovative procedure may be in the form of the adoption of an existing drug for a new purpose or changing the mixture of drugs and their dosages to adapt them to a different medical problem. In surgery, it may be in the form of a modification of an existing technique (usually accompanied by modifications of the devices being used) for application to a new use. In treatment areas that do not depend on drugs or devices (e.g., psychotherapy) or in which drugs and devices are used but are not crucial to the innovation (e.g., primary care), it may be an innovative interpretation of the existing knowledge (e.g., the multiple schools of psychotherapy which have sprung up or the new specialty of "family practice").

Increasingly, innovations in procedures arise in academic or academic-associated centers, where physical and professional resources are readily available; a research, innovation-seeking atmosphere is encouraged; and contacts with others in the field extend not only nationally, but also globally. Innovators in such settings know how to present the innovations in a manner that will be technically acceptable, and they also have the prestige that gives them access to professional meetings and journals to publicize their results. Their presentations and publications not only diffuse the innovation to a wider audience, but, more importantly, begin to legitimize it. Depending on the claimed innovation's nature, usually defined in terms of how it might revolutionize or at least substantially affect the related area of medical or surgical practice, other academic centers will begin to pursue the innovation as well.

At this point, several Federal agencies may enter the picture. The National Institutes of Health (NIH) may provide support for the innovator and researchers in other health centers in the form of randomized clinical trials (RCTs), most likely conducted in some of the clinical research centers funded by NIH. A new use for a drug, invention of a new device, or modification of an existing device requires FDA approval. Investigational new drug or device uses approved by FDA for limited testing are increasingly given to the same centers which NIH supports as clinical research centers (or at least to the health institutions in which these designated centers are located).

Sooner or later, the Health Care Financing Administration (HCFA) may receive a request for reimbursement of the procedure and will give great weight to NIH clinical trials for evidence of safety and efficacy. Meanwhile, however, FDA must make a determination of safety and efficacy for market clearance of the drug or device under review. FDA will often have to make its decision long before NIH reaches a decision and terminates funding for the clinical trials. The reason is that FDA must act in a timely manner and reach its conclusion on minimal evidence, while NIH has no similar regulatory responsibilites and is more interested in the cumulative evidence. FDA's decision, moreover, especially in the case of devices, may rest on the narrow question of the efficacy and safety of the device in a particular setting, not of the entire procedure in general use. But release of the device to the general market, once premarket approval is given, tends to speed up the diffusion of the procedure which NIH may be studying. This may place more pressure on HCFA to reimburse for the procedure.

Although there are no explicit data on which to estimate the developmental costs of medical innovation, they are without a doubt very large. Commonly, the costs of the developmental phase of early clinical application have been paid by patients, usually through standard medical insurance policies. Even for procedures that have been clearly designated as experimental, reimbursement has often been provided. Thus, for example, when total hip replacement was first introduced into this country in 1971, it was still an experimental procedure; reimbursement for the procedure was nevertheless provided.

PROGRAMS AND POLICIES RELATED TO THE DEVELOPMENT, DIFFUSION, AND USE OF MEDICAL TECHNOLOGIES

The public and other organizational policies and programs related to the development, diffusion, and use of medical technologies can be broadly classified under four headings: 1) research activities, 2) regulatory responsibilities, 3) reimbursement policies, and 4) coordination of assessment activities and dissemination of information. Some of these policies and programs have been described in previous OTA reports (266,270,274, 281).

Research Activities

R&D activities are an integral part of the innovation process. Funding for the basic research which advances medical care comes primarily from NIH, with smaller but important amounts from other Government agencies, industry, and private foundations (222). The central role that basic research plays in the process of medical innovation (64) is the justification for the substantial public and private moneys invested.

The development and diffusion phases of medical innovation are also central to the innovation process, but for these phases there is relatively little formal public funding. NIH, whose primary focus is research, appears to have no systematic or comprehensive policy of support for technology development. Although NIH grants and contracts have been given to support technology development in a number of areas (e.g., the artificial heart program, cancer screening, cancer chemotherapy, and, in recent years, hemodialysis), figures to document the size of NIH investment in development are not available. The amount invested in development probably constitutes a relatively small part of the current $3.8 billion NIH budget.

In the discussion that follows, the emphasis is on public and private research related to medical technology assessment. Information derived from this research should be useful in setting policies affecting medical technologies.

Federal Research Activities

Medical care research is of two general types: 1) biomedical research, and 2) health services research. More than a dozen Federal agencies are involved in conducting biomedical research, but the 11 institutes of NIH receive approximately 70 percent of Federal funds (228). The primary sponsor of health services research is the National Center for Health Services Research (NCHSR). HCFA sponsors additional health services research, but tends to focus it on the needs of Medicare and its other operating programs. A fourth Federal agency, the National Center for Health Care Technology (NCHCT), was established in 1978 to support evaluations of health care technologies, but did not receive funding for fiscal year 1982.

NIH and each of its institutes divide their resources among extramural grants, contract research, and intramural projects initiated by scientists within NIH. The Federal agencies that support biomedical and health services research rely on a peer review system to judge proposed projects (270). The peer review system of NIH, for example, consists primarily of non-Federal scientists and lay advisors from across the Nation grouped into 130 peer review groups, advisory committees, councils, and panels (125). They provide NIH with opinions on the scientific and technical merits of grant applications and contract proposals and on program initiatives and policy issues (270).

NIH both conducts its own testing and encourages and funds medical research and testing activities in academic centers and research institutions. It has funded studies of drugs, devices, and medical and surgical procedures, though it generally does not synthesize the evidence garnered from these efforts (266). NIH is the largest single source of funds for the support of RCTs in the country. Such trials, as described in chapter 3, are a key method for obtaining information about the safety and efficacy of certain medical technologies. In 1975, NIH provided approximately $110 million for clinical trials, a figure representing 5 percent of its total budget (266). By 1979, support for clinical trials had increased to over $135 million. Tables 2 and 3 summarize NIH's support for clinical trials during fiscal year 1979.

NCHSR conducts and sponsors a wide variety of health services research. This agency has three principal responsibilities: 1) to develop information that might be used by various decisionmakers in the public and private sectors; 2) to serve as the focal point for coordination of health services

Table 2.—Number and Amount of NIH Support for Clinical Trials Active in Fiscal Year 1979 by Institute

Institute	Number of trials	Amount of support
NIH	986	$136,160,116[a]
NEI	26	8,605,609
NHLBI	20	56,523,501
NIAID	120	6,496,938
NIAMDD	67	8,240,133
NICHD	32	4,183,244
NIDR	26	1,778,699
NINCDS	40	2,660,949
NIGMS	1	225,750
NCI	654	47,445,293[a]

[a] One trial did not report amount of support.

SOURCE: NIH Inventory of Clinical Trials, 1979.

Table 3.—Amount of NIH Support for Clinical Trials Active in Fiscal Year 1979 by Institute for Type of Support

	Extramural support					
Institute	Grant	Contract[a]	Grant and contract	Total	Intramural support[b]	Total amount of support
NIH	$47,304,588[c]	$75,738,768	$1,954,960	$124,998,316	$11,161,800	$136,160,116[c]
NEI	3,141,547	5,378,262	—	8,519,809	85,800	8,605,609
NHLBI	4,006,736	50,933,477	159,788	55,100,001	1,423,500	56,523,501
NIAID	2,435,341	3,827,597	—	6,262,938	234,000	6,496,938
NIAMDD	1,927,658	5,226,975	—	7,154,633	1,085,500	8,240,133
NICHD	3,074,448	556,296	—	3,630,744	552,500	4,183,244
NIDR	221,977	557,672	—	779,649	999,050	1,778,699
NINCDS	1,786,449	439,000	—	2,225,449	435,500	2,660,949
NIGMS	225,750	—	—	225,750	—	225,750
NCI	30,484,682[c]	8,819,489	1,795,172	41,099,353	6,345,950	47,445,293[c]

[a] Contract includes interagency agreements without intramural support.
[b] Intramural support includes intramural support in combination with interagency agreements.
[c] One trial did not report amount of support.

SOURCE: NIH Inventory of Clinical Trials, 1979.

research within the Public Health Service (PHS); and 3) to ensure that results from its research, evaluation, and demonstration activities are disseminated rapidly and in a form which is usable (290).

NCHCT, though not funded for fiscal year 1982, was required as part of its 1978 legislative mandate to undertake and support comprehensive assessments of health care technologies, including analyses of their safety, efficacy, and economic, social, and ethical implications. NCHCT had its own extramural program for awarding grants and contracts for assessments, research, demonstration, and evaluations in the field of health care technology (90,112).

In addition to responding to Medicare coverage questions (see discussion below), NCHCT identified priority technologies for "focused" or "full" assessments. NCHCT selected technologies for these assessments through the advice of the Secretary of Health and Human Services, its National Council, and others. Full assessments were comprehensive, integrated analyses of a technology's safety, efficacy, and effectiveness, and any social, ethical, or economic implications. Such assessments usually involved commissioning an overview paper, establishing a Federal planning group, establishing a full planning group, and convening a conference. Conferences were held on such topics as coronary artery bypass graft surgery, dental radiography, cesarean section, and electronic fetal monitoring (110,111).

HCFA's research objectives and priorities are defined by the information needs of HCFA operating programs. HCFA's research and demonstration mission is to improve the operating effectiveness of the Medicare and Medicaid programs. The agency's Office of Research and Demonstrations is currently conducting over 200 intramural and extramural projects on reimbursement issues, coverage eligibility, and management alternatives to present Federal programs, as well as on the impact of HCFA programs on health care costs, program expenditures, access to services, health care providers, and the health care industry.

One focus of HCFA's Office of Research and Demonstrations is on data acquisition and data management systems. Over the past 3 years, grants have been awarded to develop "integrated data demonstration" systems in a number of States, including Iowa, Maine, Massachusetts, Minnesota, Missouri, New York, South Carolina, and Vermont.* Each grantee proposes to develop a central data base—often conceived of as a "clearinghouse" or "data broker"—in which numerous types of billing and discharge abstract data, and sometimes other types of data, will be collected and linked. Although the eventual goal is statewide implementation of a system that collects data on all patients, most grantees propose initial implementation in only a few pilot or test hospitals (32,105,106,107). The Office of Research and Demonstrations has several publications for disseminating research and demonstration findings, which are also available through the National Technical Information Service.

---

*Recent decisions by HCFA have resulted in continued funding of the demonstrations in South Carolina, Maine, Vermont, and Missouri. Demonstrations in Massachusetts and Iowa were up for renewal later in fiscal year 1981.

Research and assessment issues are often initially identified by HCFA in the form of reimbursement coverage questions: Is the test, procedure, or treatment regimen provided to a specific patient "reasonable and necessary?" (305). Coverage questions originate when a bill is submitted to the Medicare fiscal intermediary, whose medical director is to determine whether an "unusual" medical event has occurred. In order to make a determination, the medical director checks through a "buddy system," contacting physicians in the local community or recognized experts in the medical field involved. If he or she decides that the intervention was appropriate, the bill will be paid, and this limited review will be the only "technology assessment" the procedure will undergo. If some question remains regarding the intervention's appropriateness, however, the bill will be forwarded to the HCFA regional office, and that office will investigate in much the same manner as the intermediary's medical director. If the intervention is accepted in the region's area, the bill will likely be paid. (Thus, the various regional offices may reach different conclusions on similar coverage questions.) In the event that there is still uncertainty, however, the regional office will pass the question along to HCFA's central office.

Until early 1980, HCFA's procedures for making coverage decisions were highly informal. The staff of the Office of Coverage Policy, often with assistance from the Health Standards and Quality Bureau, would review the issue, consult experts in the field with whom they were acquainted, and come to a decision. Although a formal agreement between HCFA and PHS had existed since around 1966 (407), a somewhat more formal process involving a panel of physicians within HCFA and from NCHCT was established in early 1980. When HCFA decided that a procedure involved a question of national importance, a request for a technology assessment was sent to NCHCT (305). Usually, such a request asked NCHCT to determine the safety and efficacy of a particular technology and to recommend whether HCFA should reimburse. Inquiries from HCFA covered the full spectrum of medical practice, ranging from the appropriateness of continued coverage for highly questionable or obsolete to medical technologies questions of reimbursement for new or investigative medical technologies. (Reimbursement policies have profound effects on the adoption and use of medical technologies and are discussed in the section on reimbursement below.)

Private Sector Research Activities

Increasingly, the private sector is involved in evaluating medical technologies, especially to determine their safety and efficacy. Manufacturers of drugs and devices initiate research and are required by FDA to conduct tests for premarket approval of their products. Large private clinics have often led the way in finding effective and efficient applications. For example, the Cleveland and Mayo Clinics were particularly active in the early evaluative efforts of the computed tomography (CT) scanner. The Cleveland Clinic has traditionally supported strong research and assessment programs in cardiovascular diseases, including an artificial heart development program. The Mayo Clinic, long recognized for its contributions to biomedical research, supports a methodologically sophisticated cadre of assessors and recently established a health care studies unit to examine problems in hospital utilization and delivery of medical services in rural areas. The health care studies unit began with $12 million in NIH grants in 1975. By 1980, its total budget approached $41 million, much of it private foundation money. Most of the unit's research can be classified as nonrandomized in design, often relying on careful recordkeeping (11,37,347).

Other research activities have been undertaken by professional associations. The American Academy of Pediatrics has developed recommendations concerning immunization practices. The American Public Health Association periodically compiles a list of effective preventive and therapeutic procedures for infectious diseases (266). The American Hospital Association and the American College of Radiology have been involved in similar activities (306).

There is also ECRI (formerly the Emergency Care Research Institute), a nonprofit organization primarily involved in comparative product evaluations of diagnostic and therapeutic devices and hospital equipment and supplies. ECRI provides a type of "Consumer Report" service for hospital administrators, which gives ratings to comparable medical technologies based on performance, safety, ease of use, and cost effectiveness. Its emphasis on the larger economic, social, and ethical issues surrounding health care technologies has recently been expanded. Further, ECRI maintains a computerized health devices data base on over 6,000 categories of devices and hospital equipment (251).

Some health insurance companies and nonprofit organizations also provide funds for research and technology assessment activities. For example, Blue Cross of Massachusetts has funded the clinical cost of an RCT comparing CT scanning, radionuclides, and ultrasound for the diagnosis of adrenal tumors, pancreatic diseases, and metastatic tumors of the liver (38). The studies were carried out at the Peter Bent Brigham Hospital in Boston and at the Johns Hopkins Hospital in Baltimore, and the cost of analysis was paid by the American Cancer Society. Blue Shield of California, along with the Multiple Sclerosis Foundation, is exploring the feasibility of similar collaborations in a clinical trial of plasmapheresis as a treatment for multiple sclerosis (38).

Approximately 15 years ago, medical insurance carriers became concerned about reimbursing new therapies that were still in the experimental phase. In 1966, therefore, Blue Shield of California established its Medical Policy Committee (primarily composed of physicians but also including members of the public and representatives of the California Podiatry Association) to assess the scope and limits of Blue Shield's standard medical insurance policies with respect to new diagnostic and therapeutic procedures (39).

In 1975, in addition to evaluating new procedures, services, and technologies, California Blue Shield's Medical Policy Committee began to identify obsolete procedures (155). This function was subsequently promoted by the Blue Cross and Blue Shield Associations under the Medical Necessity Project.

One outgrowth of that project has been the Clinical Efficacy Assessment Project (CEAP), undertaken by the American College of Physicians with funding from the Hartford Foundation. CEAP's specific objectives are to: 1) identify and evaluate technologies that are only partially effective, 2) disseminate evaluative information of potential use to health care providers and third-party payers, 3) evaluate technologies that are more efficacious in one setting than in another, 4) obtain such measures as cost benefit/cost effectiveness and marginal utility, and 5) discover how physicians decide about technology use. CEAP's mandate at this time does not include the investigation of new or emerging technologies. For the foreseeable future, therefore, CEAP will review tests and procedures that are in current use (4).

Another important initiative of California Blue Shield's Medical Policy Committee, undertaken in the interests of cost containment, was the Ambulatory Surgery Project. In 1976-77, it identified more than 700 surgical and diagnostic procedures that could normally be performed in an ambulatory setting without admission to a hospital (38).

Regulatory Responsibilities

To regulate effectively, the Federal Government must obtain adequate information in a timely manner. One major regulator, FDA, requires manufacturers of drugs and some medical devices to submit information about their products which is gathered according to approved research protocols. Other regulatory mechanisms, including local health planning agencies, are in need of information but have no particular means to obtain it.

The major Federal health regulatory activities are described briefly below. The reader will note that one primary effect that regulation tends to have on medical technologies is to constrain their development, diffusion, and in some instances, their use.

Federal Regulation of Drugs

FDA becomes officially involved in the development process for a new drug when the drug's "sponsor" (e.g., manufacturer) files a "notice of claimed investigational exemption for a new drug" (IND) for FDA's permission to test the drug in humans. If FDA approves the sponsor's IND, the sponsor may proceed with clinical testing. There are three phases in the clinical investigation of a new drug, and each phase must have been preceded by specified animal tests. (Test requirements for contraceptives are more stringent than the requirements set forth below for other drugs.)

Phase I studies are investigations of a new drug's clinical pharmacology to determine levels of tolerance (toxicity), followed by early dose-ranging studies for safety (and, in some cases, efficacy) in selected patients. The total number of both healthy volunteers and patients, which varies with the drug, ranges from 20 to 50. If the drug is found to be safe, the manufacturer can proceed to the next phase of testing. Phase I studies must be preceded by 2- to 4-week studies in two animal species.

Phase II studies, designed to demonstrate effectiveness and relative safety of a new drug, are carried out on 100 to 200 patients under controlled conditions. If the drug's therapeutic value is demonstrated and there are no serious toxic effects, the manufacturer can proceed to the next phase of testing. Phase II studies must be preceded by 90-day studies in two animal species.

Phase III studies are expanded controlled and uncontrolled clinical trials, involving 500 to 3,000 patients in usual medical care settings (clinics, private practice, hospitals). After completing clinical testing under IND, the sponsor of the drug may submit to FDA a "new drug application" (NDA). An NDA is a request for FDA's permission to market the drug. At least two well-controlled clinical trials, accompanied by complete case records for each patient, are usually required for FDA's approval of an NDA. Chronic animal toxicity studies (1-year dog, 18-month mouse, and 2-year rat studies) must be completed by the time of the NDA submission. If FDA finds the effectiveness and toxicity data acceptable, it approves the NDA. Since 1962, FDA has reviewed over 13,500 applications for INDs and has approved about 1,000 NDAs (154).

Once a drug is on the U.S. market, FDA has little control over its use or evaluation (274). Processes for collecting information on the safety (rare adverse reactions, long-term effects) and on the indications for use of drugs on the market are very limited and for the most part voluntary. In recent years, there has been increasing discussion in the United States about relying more on postmarketing controls on drugs and relaxing the premarketing controls somewhat.*

Federal Regulation of Medical Devices

The 1976 Medical Device Amendments to the Food, Drug, and Cosmetic Act greatly expanded FDA's role in regulating the safety and efficacy of medical devices. Prior to the amendments, FDA had classified devices such as soft contact lenses, pregnancy test kits, intrauterine devices, nylon sutures, and hemostats as "drugs" (359). In 1969, the U.S. Supreme Court ruled that this move was justified since Congress intended the public to be protected from unsafe and ineffective devices (299). The Medical Device Amendments of 1976 established a three-tiered system of controls: Class I, General Controls; Class II, Performance Standards; and Class III, Premarket Approval. Each device is required to be classified on the basis of the level of regulation needed to ensure its safety and efficacy.

Class I devices are low-risk devices that are not used to support or sustain human health, and these are subject primarily to the Food, Drug, and Cosmetic Act's basic prohibition against misbranding and adulteration. Although Class I controls apply to accuracy in labeling and the sanitation and physical integrity of low-risk medical devices, all devices must meet these minimum standards. FDA also has the power to ban any device, regardless of classification, which presents a substantial deception or an unreasonable and substantial risk of illness or injury that is not correctable by labeling.

Class II devices are those for which general controls alone are judged insufficient and about which sufficient information exists or could be developed to establish performance standards. FDA is authorized under the 1976 amendments to develop and establish performance standards.

---

*This topic is explored at greater length in OTA's report *Postmarketing Surveillance of Prescription Drugs* (281).

Class III devices are those devices that are life-sustaining, life-supporting, implanted, or present a potential unreasonable risk of illness or injury, and for which general controls or performance standards may not provide reasonable assurance of the device's safety and efficacy or for which performance standards cannot be developed. Class III controls are comparable to the premarket approval process for drugs.* Any device which was classified as a "drug" before 1976 is automatically assigned to Class III unless reclassified. Any device developed after 1976 which is not judged by FDA to be "substantially equivalent" to a preamendment device in Class I or Class II will also be assigned to Class III and require a premarket approval application. In the first 4 years after implementation of the 1976 amendments, about 98 percent of the listed devices in the 10,540 premarket notifications received were declared "substantially equivalent" to a preamendment Class I or Class II device (270).

The 1976 amendments require any distributor of a device intended to be marketed for the first time to file a notice with FDA at least 90 days in advance to permit the agency to decide whether the device needs premarket approval to assure safety and efficacy. FDA permits earlier distribution if it concludes and notifies the distributor that premarket approval is not required. If the 90 days pass without comment from FDA, marketing can begin. In 1981, FDA estimated that 2,300 premarket notifications would be reviewed.

Industry often uses FDA approval to advantage for its marketing strategy. All results of clinical investigations will ultimately be included in a package insert, product data sheet, or physician's brochure, which are FDA-approved generators of promotional claims (300).

Other Regulatory Activities

Concerns about premature diffusion of the more expensive devices and other capital investments led to the enactment of three overlapping Federal programs: 1) section 1122 review, 2) the 1974 National Health Planning and Resources Development Act, and 3) State certificate-of-need (CON) laws. Concerns about the utilization of health care services by beneficiaries of the Medicare and Medicaid programs led to the development of Professional Standards Review Organizations (PSROs).

The first State CON law was enacted by New York in 1964. This law, subsequently followed by similar laws in other individual States, empowered State planning agencies to deny reimbursement to hospitals for large capital expenditures unless the planning agency found a "need" for the service to be provided. In 1972, section 1122 of the Social Security Act similarly authorized the Medicare and Medicaid programs to withhold funding for depreciation, interest, and return on equity capital for certain investments found not necessary by a health planning agency. State CON laws and section 1122 review, in effect, constitute a franchising process for potential adopters of expensive medical technologies.

Section 1122 applies to investments of more than some specified amount (initially $100,000) and covers changes in beds and services that are provided by certain health care facilities, such as ambulatory surgical facilities. Private physicians' offices are explicitly exempted. In 1977, 37 States had contracted with the Department of Health and Human Services (DHHS)* to conduct section 1122 reviews.

The National Health Planning and Resources Development Act of 1974 required States to pass CON laws by 1983 as a condition of future Federal funding under the Public Health Services Act, the Community Mental Health Centers Act, and the Alcohol Abuse and Alcoholism Act. The original act applied to the same facilities covered by section 1122 review. However, 1979 amendments to the act exempted health maintenance organizations (HMOs) from having to secure a CON for inpatient investments.

PSROs, enacted into law in 1972, are areawide groupings of practicing physicians responsible for reviewing services provided and paid for by Medicare and Medicaid. The purpose of their review is to help assure that these services are: 1) medically necessary, 2) of a quality that meets locally determined professional standards, and 3) provided at the most economical level consistent with quality of care. However, the primary operational mission of PSROs has been to constrain excessive utilization of health care services which is fueled by the reimbursement incentives discussed in the next section of this chapter.

---

*The 1976 amendments also allow FDA to permit developing and marketing approval of a Class III device under a "product development protocol," where FDA and manufacturer agree in advance on a plan for the development, testing, and release of the device. This approach has not been implemented.

*Then the Department of Health, Education, and Welfare.

Other regulatory-type mechanisms that have been instituted because of the high demand generated by third-party payment include State hospital rate-setting programs, increasing Medicare deductibles, setting low reimbursement levels for the Medicaid population, decreasing Medicaid benefits, and raising Medicaid eligibility requirements. All of these affect the diffusion and level of use of medical technologies.

Reimbursement Policies

Reimbursement policies have profound effects on the adoption and use of medical technologies. Reimbursement also influences the innovation process, especially for medical and surgical procedures. With the increasing costs of medical care continuing to cause concern, reimbursement policy is becoming even more important. Informed coverage decisions may require even more detailed information concerning medical technologies than regulatory decisions. Whereas regulatory decisions tend to be more of a "go," "no go" nature, reimbursement decisions are, or at least could be, more related to appropriate use of technologies, a much finer distinction.

The growth in third-party coverage of medical care is seen as a major cause of the excessive adoption and use of many medical technologies (142, 331). It is important to note, however, that factors other than reimbursement policy contribute to the overall tendency to adopt and use medical technologies at excessive levels. Such factors include competition among hospitals to achieve quality and prestige to attract patients and physicians, public demand for sophisticated technologies, increasing specialization within medicine, physicians' desires to do as much as possible for their patients, uncertainties related to what constitutes appropriate use, and the defensive overutilization of medical tests and procedures because of the threat of malpractice suits.

Variations of Reimbursement Mechanisms

There are two basic forms of payment mechanisms in the U.S. medical care delivery system: cost-based and charge-based (305). Government programs, primarily the Medicare and Medicaid programs, were developed to "buy into" what was then perceived as a market pricing system. The statutes enacted in 1965 established the principle that the Government purchaser would pay institutional providers the costs of services to patients. Physicians were to be paid their "usual, customary, and reasonable" fees. The original assumption was that the Government was buying at the margin and would not affect the average costs of the system. Subsequently, however, it came to be recognized that Government purchases of medical services were sufficiently large to affect purchase price and costs. Thus, the 1972 amendments to the Social Security Act placed limits on the amount that would be paid by Medicare to both institutional providers and physicians. Rather than being related to efficiency, these cost limits reflected rates of increases in charges over time.

In the "private" sector of the medical care market, there are two widely used mechanisms to set reimbursement levels. One, the cost-based Blue Cross/Blue Shield reimbursement system is in many ways similar to the Medicare program. Hospitals are reimbursed the "reasonable" cost of providing care to patients, and physicians are paid "reasonable" fees. The second mechanism is payment for billed charges and is used by some Blue Cross/Blue Shield plans and in all contracts established between patients and other insurers to pay the bills generated by the patient. Under this approach, all or some of the charges of hospitals and other medical providers are paid through insurers, unless there are copayments and deductibles which are paid by the patient. Patients not covered by Government or other insurers are responsible for their own bills. Billed charges are more like a market mechanism, except that demand is not directly affected by the income or wealth of the patient.

A third payment mechanism, not very widespread, is capitation, whereby a fixed amount is paid for each patient per time period, regardless of the health services provided. The capitation method generally involves the integration of financing and the delivery of services, thus placing the provider of medical care at financial risk.

Influence of Reimbursement on the Development, Diffusion, and Use of Medical Technologies

When coverage for new and experimental medical and surgical procedures has been offered from the outset, a high level of reimbursement has been justified on the basis of the special skills and large amounts of professional time required, and perhaps on the basis of increased risk. When such procedures have become routine, requiring less time and skill and posing lesser risks, however, professional fees have usually increased rather than fallen (316).

Several examples have been provided by Blue Shield of California (39). Phakoemulsification of the crystalline lens, introduced as an alternative to lens extraction for cataract, is—once learned—shorter and no more complex than standard lens extraction, yet surgeons initially charged 25 to 30 percent more for the new procedure than for the older, more costly procedure. In this instance, California Blue Shield's Medical Policy Committee disallowed the increase. Another example is the flexible fiberoptic endoscope. Although this new instrument is easier to use than the standard rigid instrument, physicians introducing the new procedure charged 25 percent more. A third example is arthroscopic menisectomy for torn knee cartilage. Orthopedic surgeons introducing this procedure wished to charge the full fee for the standard open arthrotomy and an additional fee for the arthroscopy. In this instance, Blue Shield of California agreed to pay the full arthrotomy fee and an additional 50 percent of the arthroscopy fee. The rationale for Blue Shield's concession was that the performance of the simpler procedure might eliminate the need for many days of hospitalization and laboratory tests, with a considerable net savings in total charges.

Allowing a simpler procedure to be billed as a more complex procedure has resulted in questionable increases in physicians' fees. In the example of arthroscopy of the knee, the large difference in allowable charges when an operative procedure is added to a diagnostic procedure offers a strong invitation to remove some tissue during arthroscopy. During the diagnostic examination of the knee, a small piece of redundant synovial membrane may be seen—a finding of no great importance. Removing a piece of this tissue makes the procedure a "synovectomy," for which the customary charge is $1,300, rather than simply a diagnostic arthroscopy, for which the customary charge is only $500. The above scenario presents a situation that may be reasonably justified medically, but, even interpreted generously, there is a clear fiscal invitation to perform a procedure that is more, rather than less, complex.

There is also a much more serious consequence of the manner in which charges are submitted for experimental procedures. With increasing scrutiny by third-party payers of bills submitted for new procedures and more than occasional denial of payment for such bills, there is a strong incentive for physicians to request payment for a standard procedure rather than a new one. This practice is also encouraged by the fact that new procedures often do not have a procedure code number, by which most bills are processed. Requesting payment for a standard procedure may simply reflect an honest effort to use whatever code number seems most nearly to approximate the procedure actually performed. The net result, however, is that the identity of the new procedure may be concealed, and the fact that an experiment has been carried out may not emerge.

In bills submitted to Blue Shield of California, there is an approximately 15-percent error rate in the coding of all procedures (39). The medical director estimates that 1 percent of the errors involve the use of existing codes for procedures to which new codes have not been assigned. Any innovation that falls outside of "accepted medical practice" is particularly vulnerable to being mislabeled. Because it is difficult to define exactly what constitutes accepted medical practice, the new procedures that have the best chance of being reimbursed are the ones that deviate the least from existing procedures which are already being reimbursed. The Federal Government, for example, has traditionally favored coverage of new technologies perceived to be modifications of existing interventions (270). The incentives, therefore, are toward the development of parallel procedures or extensions of existing technologies.

For procedures that deviate substantially from accepted medical practice, the reimbursement system may require considerable testing for safety, efficacy, and costs to determine if they offer sufficient contributions to compensate for their deviation from standard medical practice. These circumstances have several implications. First, when procedures remain outside the coverable range, they may also suffer the fate of anonymity, neglect, lack of funding, or underutilization. An obvious example is the traditional exclusion from most insurance plans of preventive medical care, most notably screening services. Second, the scrutiny of radical innovations rather than of incremental improvements may be misplaced to the extent that the growth in medical expenditures is the primary reason for such scrutiny. The collective expense of small tests and procedures is arguably far greater than that of a few "big ticket" technologies (249). Third, if radical innovations have the most difficulty in receiving favorable coverage decisions, innovators might be inclined to pursue less radical but more easily accepted innovations. This is a difficult hypothesis to test, as radical innovations have less chance of commercial success than minor innovations; but once radical innovations penetrate the market, the magnitude of their commercial success is greater than for minor innovations. Fourth, a technology-by-technology approach to coverage decisions, with priorities determined by how radically each technology differs from existing ones, may invite those seeking payment for the use of new technologies to submit their claims for payment under the guise of accepted procedures.

Under either cost reimbursement or charge payment, third-party payments generally are intended to cover the full costs of new technologies, including purchase, maintenance, or operation of equipment; the leasing of equipment; the costs of drugs; or the facilities and equipment needed for a procedure (19). One would expect that greater adoption of technologies would occur under these relatively price-independent conditions than would occur under a more price-sensitive system. Cromwell, et al.'s (75), interstate analysis found that the hospital's percentage of revenues from third parties was significantly and positively related to the hospital's adoption of expensive technology. Russell (331) found that the adoption of cobalt therapy and electroencephalography occurred faster when the level of insurance coverage was higher and proceeded more rapidly as that level grew. She also found that a greater contribution of hospital costs by Medicare was associated with increased adoption of cobalt therapy, intensive care beds, and diagnostic radioisotopes. And Willems (392) concluded that open-heart surgery spread more quickly in areas with faster growth in insurance coverage.

Third-party reimbursement can indirectly affect the adoption of technology by changing the availability of financial capital to potential adopters. A prominent example is the Medicare program, which reimburses institutional providers for capital as well as operating costs. Medicare payment for allowable capital costs such as depreciation and interest provides a source of internally generated funds (28). Third-party coverage, especially by Medicare and Medicaid, has also reduced hospitals' risks of bad debts, thereby improving their standing as credit risks to private lenders. Other changes in governmental programs, such as the Hill-Burton program for funding medical facility construction and modernization, as well as various tax-exempt bond programs, have also affected the sources of financial capital.

In addition to affecting the adoption of technologies, the extent of third-party coverage would be expected to affect the use of technologies. Data on the use of specific technologies are generally lacking, however. Cromwell, et al. (75), found that many hospital technologies are underutilized after being adopted. Nonprofit hospitals in the Boston area were using automated analyzers and patient monitors (and, in teaching hospitals, diagnostic X-rays) at only half of capacity. Willems (392) considers such underuse as presumptive evidence of the hospitals' overinvestment in new equipment.

It is not clear how this relatively price-independent adoption of medical technologies is used by medical care providers to compete with one another. As summarized by Banta, et al. (19):

> Studies of hospitals found no definite relationship between measures of competition and adoption. The situation is complex, because the characteristics of the market may relate not only to competitiveness, but also to the availability and sharing of information and to local standards of practice. The evidence conflicts, depending on the characteristic used and the technology studied. Russell (331) found that concentration of market power among a few large hospitals did not appear to influence the adoption of three common and two prestige technologies, but that hospitals in more concentrated markets were less likely to adopt open-heart surgery. Prior adoption in a locality reportedly speeded the adoption of intensive care units and electroencephalographs, but not diagnostic radioisotopes, open-heart surgery, renal dialysis, cobalt therapy, and computers (75,331). In urban areas, greater adoption of radioisotopes and electronic data processing occurred where there were many hospitals per capita, the hospitals were of similar size, and they were close to other hospitals (212,301).
>
> Different patterns have also been observed between adoption and the number of physicians per capita. Facing a low physician-population ratio, hospitals may compete for physicians through technology adoption. On the other hand, fewer physicians may exert less pressure for adoption. The adoption of CT scanners and radioisotopes appeared unrelated to the physician-population ratio (301,392). However, greater adoption of intensive care units, open-heart surgery, cobalt therapy, and renal dialysis occurred among States with higher ratios (75).

Thus, even though current payment mechanisms for medical care services can lead to excessive adoption of medical technologies, there are still constraining factors which make it clear that cost is not the only factor which influences adoption.

Coordination Efforts and Dissemination of Information

Federal Activities

The Technology Coordinating Committee of DHHS served as an interagency forum for the identification and discussion of problems and issues associated with health care technologies (110,111,112). This committee, previously chaired by the Director of NCHCT, fostered information exchange and interagency cooperation on health care technology matters and has served as the department's principal mechanism for joint action on appropriate issues. Now that NCHCT is no longer funded, DHHS is studying whether to keep the Technology Coordinating Committee and, if so, how to organize it.

NCHSR has responsibility for coordinating health services research within agencies of PHS. To coordinate this research, NCHSR chairs the PHS Health Services Research Coordinating Committee, which includes representatives from each of the PHS agencies.* NCHSR also meets regularly with HCFA to review research priorities and to determine how each organization's research activities might contribute to the other's programs. In fiscal years 1980 and 1981, NCHSR and HCFA produced a joint health services research strategy and budget (113).

NCHSR also disseminates research results to relevant Government agencies, the research community, and other interested parties by means of publications, press releases, conferences, and workshops. In 1978, the legislation authorizing NCHSR was modified to require that at least $1 million or 5 percent of its budget, whichever is less, be used for dissemination activities. In response, NCHSR established a User Liaison Program, aimed at providing substantive assistance to non-Federal health care leaders concerned with critical policy issues and operational problems in the organization, administration, regulation, and delivery of health care services at the State and local level. In 1979, NCHSR's User Liaison Program conducted nine workshops that were attended by users of health services research such as State legislators, executives of State health agencies, leaders of both the insurance industry and hospital sector, and city health officials (113).

---

*NCHSR's Health Care Technology Study Section, which served as the scientific peer review committee in the grants review process for NCHSR and NCHCT, provided an additional formal coordination link.

The Office for Medical Applications of Research (OMAR), established in the NIH Director's Office in 1978, monitors, facilitates, and evaluates NIH research, technology assessment, and technology transfer activities. As noted in chapter 5, OMAR coordinates NIH consensus development conferences. The OMAR Advisory Committee—consisting of representatives from the bureaus, institutes, and divisions of NIH and from other Federal agencies, including FDA, the Centers for Disease Control (CDC), and the NCHCT (while it was funded)—assists OMAR in its planned consensus development activities. The committee also assists OMAR in the exchange of information relating to other NIH involvement in assessing biomedical technologies (228).

The consensus conference panels of NIH are composed chiefly of medical experts, although they also include members of the lay public and selected professions (e.g., clerical and legal). The technologies these panels examine may be emerging technologies or technologies in general use and may be drugs, devices, or medical, surgical, or dental procedures. (As described in ch. 5, there have been over 30 consensus conferences since the first on breast-cancer screening in September 1977. The topics and dates of all conferences from 1977 through the end of 1982 were listed in table 1.) On topics representing areas of mutual interest and concern, consensus conferences have been sponsored by NIH in conjunction with NCHCT, in collaboration with an agency outside NIH, or under the cosponsorship of two or more institutes within NIH.

An essential part of the OMAR consensus development program is the dissemination of consensus statements and supporting materials to practicing physicians and others in the health care system, the biomedical research community, and the public. It is hoped that by supplying medical practitioners with critiques of complex medical technologies, dissemination of these reports will contribute to an improvement in the quality of medical practice. OMAR has compiled a mailing list of over 21,000 names. Consensus materials and information have been published in three major American medical journals (*Journal of the American Medical Association, New England Journal of Medicine,* and *Annals of Internal Medicine),* as well as in State medical society periodicals and the general press (229,287).

NCHCT initiated the proposed development of a Clinical Data Acquisition Plan, a conjunctive effort with both public and private group support to develop a model method for collecting clinical data on emerging technologies. Under the model process, third-party payers would provide interim reimbursement for appropriate technologies and related services to those providers who agreed to submit certain prescribed data (110,372).

As called for by its 1978 authorizing legislation, NCHCT also compiled an annual "emerging technology list." All DHHS and other relevant agencies submitted a list of candidate technologies, including background information and a preliminary assessment of each. Although the "emerging technology list" was intended only to identify emerging technologies, not necessarily to assess them, it came under increasing attack by industry as a threat to innovation. The mere appearance of a technology on the list, it was argued, increased managerial reluctance to develop the technology because it created additional uncertainty in further marketability (111).

Other Processes

Apart from those mechanisms involving Federal agency interaction, various other mechanisms by which medical technology information is distributed include: 1) the public media, 2) the mail, 3) advertising, 4) personal contacts, 5) the educational process, and 6) libraries and other types of information providers, including Federal information centers and private for-profit and not-for-profit organizations. While the relative importance of each is arguable, the appropriateness of individual mechanisms may partly depend on whether the information is to be used in conducting an assessment of a medical technology or is to be used in conveying the results of a medical technology assessment.

The *popular print media,* including daily and weekly newspapers and journals, are a primary channel of information on health, including medical technology assessment, for the general public. At times, they also serve a similar function for physicians, nonphysician health professionals, legislators, and others in the health field. Joining the mass circulation publications are an increasing number of biweeklies and minimagazines serving special interests, regions, and even localities (370). Many of the mass circulation publications employ a trained science/health columnist at regular or occasional intervals. In addition, news reports of immediate happenings in health, health advice columns, retrospective analyses of technologies, and more often, predictions about the future of new technologies are found in all forms of print. Indeed, some popular publications are devoted exclusively to health and/or specific aspects of health.

The diversity in the print media is paralleled in *radio* and *television*. Some networks and/or stations, especially publicly owned or operated outlets, occasionally explore a medical technology in depth. One can question whether the 5-minute-or-so news programs provide health information of any real value, but such programs usually carry spot announcements about health fairs and the need for their listeners to take advantage of technologies, such as immunizations and high blood pressure medications. Health fairs also supply information about medical technologies, as well as how and where professional assistance can be obtained.

Mailings are a common mechanism for disseminating both unsolicited and solicited information for and about medical technology assessment. Unsolicited information is that which is received without having been requested or paid for by membership or subscription. Among the materials available by this mechanism are direct mailings about medical technologies (e.g., newsletters from drug companies) and advertisements from product distributors. Unsolicited information through the mails is an important source of health information for both lay people and health professionals. Most Federal agencies dealing with medical technology use the mails for sending literature in response to direct requests or to people on their mailing lists.

Advertising of drugs, and to a much lesser extent of other medical technologies, is prominent in all the popular media. The large budgets that most pharmaceutical companies allocate for this purpose seem presumptive evidence that there is a market for this source of information among the general public. Advertising is termed education by the companies involved, especially when the target audience is physicians and other health professionals. Pharmaceutical manufacturers spend several hundred million dollars a year in advertising their products to the medical profession in professional journals, at professional meetings, and through their representatives ("detail men"). These expenditures would not be likely to continue if they did not bear results (65).

Recently, some drug companies have been supplying hospitals and other medical facilities with video cassettes that contain information on various aspects of health care including medical technologies. They also are producing closed circuit television programs on similar topics to be received at medical facilities that have video tape receivers (214).

For a discussion of the ways in which physicians keep informed about, and are influenced by, new developments of medical technologies, the reader is also referred to an earlier section of this chapter concerning the diffusion of medical technologies.

CONCLUSION

This chapter has discussed the innovation and diffusion of medical technologies, especially medical and surgical procedures, and policies that affect the innovation and diffusion process. Many of the points raised in this chapter are discussed at greater length in appendix D.

The innovation process is complex and not well understood, but is certainly important to an assessment strategy. Most regulation is intended to substitute for an imperfect market. Government has adopted a general sense of public responsibility by seeking to ensure that unsafe and inefficacious drugs and medical devices not be allowed on the market. The thrust of nearly all FDA regulations is to require manufacturers of new drugs and certain medical devices to test their products for safety and efficacy according to approved protocols. FDA then synthesizes this information, decides whether to approve the marketing of the technology, and then regulates the labeling process. Thus, FDA is involved with all stages of medical technology assessment as defined in chapter 1 (i.e., identification, testing, synthesis, and dissemination). However, FDA's involvement is limited to certain types of technologies—emerging and new drugs and devices. Also, FDA's activities are generally limited to assessments of safety and efficacy; cost, cost effectiveness, and other social/ethical effects are generally not explicitly considered.*

The information requirements of FDA tend to slow the innovation and diffusion of certain medical technologies. Industry, especially the medical devices industry, is concerned that FDA's information requirements unnecessarily threaten innovation. A major problem in analyzing industry's concerns, however, is the difficulty of determining the costs and benefits of testing requirements.

Reimbursement policies also distort the innovation process. In general, it appears that the wide availability of medical insurance contributes to the overadoption and use of many medical technologies. In many cases, the lack of technology assessment information at the point of reimbursement tends to speed up the diffusion process. As suggested earlier, however, the diffusion of truly innovative technologies that fall outside generally accepted medical practice may actually be discouraged by the present reimbursement system.

This dichotomy seems to be related to the lack of adequate scientific evidence of the value of new technologies. When a new technology is an add-on (i.e., when it does *not* directly substitute for an existing technology), produces more information, or for some reason captures the imagination of the medical profession, the technology tends to be accepted and even encouraged by the medical profession without substantial evidence of its value. On the other hand, radically new technologies that challenge preexisting beliefs—or in some cases merely the status quo—are less likely to be acceptable to the medical profession without very strong evidence of their worth. Since present reimbursement policy rests in large part on accepted medical practice, these more radical technologies tend not to be acceptable for reimbursement, and their innovation may thus be discouraged.

The effects of regulation and reimbursement policies on the innovation process are clearly interrelated. As new medical procedures develop, they often make use of new drugs and devices or use existing ones in modified ways. Such drugs and devices generally have to pass through FDA's regulatory process. Until these technologies are approved for marketing, regulatory review acts as a constraint on the adoption and dissemination of the procedures in which they are used. Once these accessories are released into the marketplace, however, they can act to stimulate use of procedures which are still experimental and not accepted medical practice. For example, FDA released the catheter used in percutaneous transluminal coronary angioplasty (PTCA) from investigational device status and approved its marketing for PTCA while the procedure itself was still considered by many to be experimental.

The Federal agencies responsible for medical research, regulation, and financing engage in a variety of technology assessment-like activities. Although there are increasing efforts to improve coordination between these agencies and collaboration with the private sector, these efforts currently fall short of a strategy for medical technology assessment as discussed in chapter 1. Current coordinating efforts are heavily oriented toward the question of reimbursement. One of the former NCHCT's formal responsibilities was to advise HCFA on coverage decisions. Current efforts such as NIH consensus development conferences are oriented toward determining the efficacy, safety, and appropriate use of medical and surgical procedures, but they can also help to provide information for HCFA's decisions regarding the reimbursement of new technologies. A major weakness of all these activities is that no body is charged with evaluating the economic and social/ethical effects of medical technologies. In contrast, the regulatory agencies, principally FDA, have limited roles in current coordinating efforts. Although FDA's regulatory responsibilities make FDA an important generator and repository of assessment data, there is little coordination of its functions with those of the other governmental health agencies.

---

* Sometimes, however, its review extends farther. For example, in the case of the injectable contraceptive Depo Provera, FDA based its decision to deny market approval, not only on safety criteria (i.e., its concerns over Depo Provera's cancer-causing potential), but partly on the basis that the patient population originally targeted for Depo Provera had diminished substantially as other methods of contraception and sterilization became increasingly available and accepted (193).

The next chapter presents OTA's critique of the current system for the identification and testing of medical technologies and the synthesis and dissemination of technology assessment information.

7.

Critique of the Current System

It is one thing to show a man that he is in an error, and another to put him in possession of truth.

—*John Locke*

Contents

Page
Introduction 91
Assessment of Medical Technologies 92
Critique of the Current System 93
  Identification 94
    Identification of Emerging Technologies 95
    Identification of New Technologies 96
    Identification of Existing Technologies 97
    Identification of New Applications of Existing Technologies 98
  Testing 98
  Synthesis 100
  Dissemination 101
Conclusion 102

7.

Critique of the Current System

INTRODUCTION

It has been well established that there is insufficient information regarding the costs, risks, and benefits of medical technologies. One purpose of assessing medical technologies is to produce information to help guide their appropriate use. Clearly, society wishes to promote the development and diffusion of safe and effective medical technologies. At the same time, society wishes to reduce ineffective and inefficient use of medical technologies. Finding a balance between these two goals is difficult. The complexity of our society and the mixed private/public nature of the health care system magnify the challenge of improving existing policies and processes for medical technology assessment.

Federal policies toward medical technology have developed in incremental fashion to meet rather specific goals. Only recently have the numerous policies that pertain to the development, diffusion, and use of medical technologies begun to be seen as elements of an overall system for guiding and promoting technological change. The collection of programs and activities that forms the current "system" for assessing medical technologies* has also been built-up in piecemeal fashion. Although many have realized that better information on the benefits, risks, costs, and social implications of medical technology is essential to guiding the development and use of technology without unnecessarily impeding innovation, progress toward developing a coherent system for assessing medical technologies has been slow.

The idea of a "strategy" for assessing medical technology is closely related to having a system. A strategy is in effect the underlying basis for the design and implementation of any coherent system for assessment. Any strategy will represent a compromise among competing perspectives on the goals of medical technology and on the role and format of assessments. Strategies reflect desired policy directions and ideas on how best to move in those directions. Without a conceptually clear strategy, any system of assessment will suffer from inconsistencies and unclear objectives.

The basic objective of a strategy for medical technology assessment is to ensure that technologies of public policy importance are evaluated by appropriate methods in a timely fashion without unnecessarily harming innovation. This objective must be sought despite a number of formidable difficulties. One problem is the very limited amount of money available for original evaluations of medical technology. Costs of a single randomized clinical trial (RCT) can run as high as hundreds of thousand of dollars, yet the entire societal investment in original studies of medical technology probably does not exceed $200 million at any one time.

A sound strategy for assessment must take into account the stage of development of particular medical technologies—emerging, new, or existing. If emerging technologies are assessed too early in their development, innovation may be slowed; furthermore, the information gained by assessing the technologies may be valueless, because knowledge about their modifications and eventual uses will be limited. If new or existing technologies are assessed too late, the assessments will have little effect on the technologies' diffusion or use. Assessment information must be disseminated to appropriate parties in a timely manner.

A strategy for medical technology assessment must also deal with the universe of drugs, devices, and medical and surgical procedures—diagnostic, therapeutic, and preventive. Other important classes of medical technology to be dealt with are medical care delivery systems and organizational innovations, although, as noted earlier, they fall beyond the scope of this assessment.

---

*By the existing "system" of assessing medical technologies, OTA means the current set of programs and activities which are related to assessing medical technologies. In this context, the word "system" is used in the general sense and is not intended to imply that there currently is some overall design involved.

A further objective of a strategy is to develop criteria to choose an appropriate method or methods for assessment. A strategy must permit one to determine when it is enough to know that a technology is efficacious, when it is desirable to have a formal cost-effectiveness study, and when a full-scale technology assessment with evaluation of social implications would be helpful and desirable.

A strategy should also address how the information gained from technology assesssment is utilized and by whom. And finally, it is necessary to consider how technology assessment should affect health policy.

These are some of the challenges. This report cannot deal with all of them. The section that follows discusses the development of information on which to base decisions. The chapter ends with a critique of the current technology assessment system.

ASSESSMENT OF MEDICAL TECHNOLOGIES

Assessing medical technologies is a complex process, and no simple model can be devised to outline the steps that must be taken in all circumstances. Technologies are diverse, often lending themselves to be evaluated in diverse ways. The need for information about technologies by different people at different points in time varies, as well. Nevertheless, there are a finite number of technologies and a finite number of assessment methods. The information that is required to make rational and reasonable informed choices is also finite.

The previous chapters of this report have identified the needs for assessment information and the resources available to fill those needs. The needs are defined by those called on to make decisions. For example, the Food and Drug Administration (FDA) must decide whether to allow a drug or device to be marketed; it asks whether the technology works and whether it is sufficiently safe. The National Institutes of Health (NIH), in its basic research efforts, must set priorities regarding which technologies, including especially medical and surgical procedures, it will further investigate and develop. Having been called on by Congress to synthesize what is known about medical technologies in order to assist in their transfer, NIH also asks what the appropriate conditions and standards of use are. Similarly, Professional Standards Review Organizations (PSROs) ask whether local practice patterns conform to reasonable standards of care. Those who pay for care, whether they be the Health Care Financing Administration (HCFA), Blue Cross, or individual patients, need to know whether the use of a medical technology is worth the cost. And, finally, the practicing clinicians, in consultation with individual patients, must make the final decision to use a technology. Throughout this process, the values and needs of society, the medical profession, and the patients themselves are interwoven.

A strategy for medical technology assessment must take into account what information is known, what is not known, what is needed, what can be obtained, and what the cost of obtaining it will be. Information will never be perfect, and money and time will always be limited. Thus, it is important to make judicious use of evaluation methods. Fortunately, there are means to compensate for uncertainty when important information is lacking.

No clear-cut rules seem to be possible in devising a strategy for choosing a method to assess the safety, efficacy, and effectiveness of medical technologies. Often, the method of choice will be related to both the stage of diffusion of the technology and the extent of knowledge and belief as to its risks and benefits.

RCTs, for example, tend to be appropriate when a technology's risks and benefits are not well understood, when the technology is not yet in general use, and/or when costs of the technology are very high in relation to expected benefits, and when risks are expected to be low. Under these conditions, the purpose of an RCT is to establish a cause-and-effect relationship. Reasonable candidates for RCTs would thus be new drugs, new invasive devices, new expensive equipment, and new elaborate services requiring capital expenditures (e.g., neonatal intensive care units). When risks and benefits are not well known or are not believed, randomization can be used without violating some of the ethical principles noted in chapter 3. If RCTs are used early in a technology's diffusion, nonrandomized designs, especially case-control studies, can later be used to establish the effectiveness of the technology as it diffuses into diverse settings.

However, when a technology is in widespread use, risks and benefits are either already known or are widely believed to exist, and randomization may be neither possible nor appropriate. In this case, nonrandomized designs can be used to establish relationships which can later be tested, if desired, by more rigorous methods, including randomization.

Economic analyses are similarly varied, and no one technique is applicable in all cases. However, economic information may be worthless without good safety and effectiveness information. For the user of the technology, the price is the cost. That price must somehow be compared to the perceived value of the use of technology. But for more general decisions, especially at the societal level, economic analyses are very complex, requiring both technical expertise and good judgment. A cost in one instance may be ignored or even counted as a benefit in another.

As discussed in chapter 4, decisions concerning the development and use of certain medical technologies often have profound social and ethical implications. Especially at the Federal policymaking level, these implications are important to consider, even though they cannot be precisely quantified.

Finally, informed decisions rest on the analysis of all available information. Chapter 5 discussed a number of techniques that can be used to synthesize information from research studies in a systematic manner. Additionally, group process techniques such as Delphi and nominal group process are available to assist policymakers and technically expert groups in making decisions. None of these methods for synthesizing information is perfect, but each has potential value in the development of more orderly processes for setting policies regarding medical technologies.

CRITIQUE OF THE CURRENT SYSTEM

The present system of medical technology assessment, like the medical delivery system, is pluralistic, and many of the public and private sector activities reviewed in this report were undertaken for purposes other than medical technology assessment. The diversity of activities is not necessarily a weakness. Such diversity capitalizes on the wealth of ideas and interest of many different people and organizations. Nevertheless, it makes the job of fashioning a more coherent system of assessment more difficult.

Perhaps the principal reason for the difficulties with the present system is that the main parts were developed separately over a long period of time with specific, sometimes inconsistent, goals. The existing programs and activities were not devised as elements of an overall system of technology assessment. In the case of the Food, Drug, and Cosmetic Act, amendments over the past three quarters of a century have been internally consistent, tending to build on and complement previous legislation. However, most other legislative and nongovernmental efforts that affect medical technologies have not been so well coordinated. Thus, the country has a system of physicians, hospitals, planning agencies, PSROs, health survey activities, research activities, and insurance claims networks, all of which use, dispense, regulate, evaluate, collect information on, or otherwise affect medical technologies, but which often do not complement one another's needs for technology assessment.*

---

*This criticism does not necessarily apply to the Veterans Administration's system, which OTA did not study to any appreciable degree. Nevertheless, it is known that the Veterans Administration is developing a system whereby potential investigators are informed of program needs, a research agenda is developed to satisfy those needs, and useful information generated by research is made available to those who need it (168).

To examine the extent to which the needs of an overall system for medical technology assessment are met, the programs and activities comprising the present system are discussed in the remainder of this chapter with reference to the four phases of the technology assessment process mentioned earlier: 1) the identification of technologies needing assessment, 2) the testing of technologies to develop information concerning their health and economic effects, 3) the synthesis of information, and 4) the dissemination of the information that is available.

OTA finds that the current system for evaluating medical technologies exhibits major deficiencies in each of the four phases of the assessment process. For technologies at different stages of development (i.e., emerging technologies, new technologies, existing technologies, and new applications of existing technologies), as well as for technologies classified as either drugs, devices, or surgical or medical procedures, the adequacy of the present system differs.

The existing system for identifying technologies to be assessed, except for FDA's system of identifying new drugs and devices, is unnecessarily poor. (Among the many reasons for this is the inadequacy of the synthesis of research information.) In the testing of medical technologies, many studies generate evaluative information, but the quality of such information varies widely. FDA's research requirements for new drugs and devices seem adequate for the premarket approval process, and much NIH-sponsored research has resulted in significant information for society. In other areas, however, high-quality studies are few, and most of them are not helpful in setting policy. High-quality, objective syntheses of research findings—a prerequisite for developing policy or setting medical practice standards—are rare. Many syntheses are informal, overly subjective, group-generated norms and are not based on a rigorous assessment of the scientific evidence. Although there are increasing efforts to disseminate technology assessment information, much of the information has questionable value. The excessive adoption, diffusion, and use of some medical technologies indicate a need for improved dissemination efforts.

In the expanded critique that follows, special attention is paid to the identification of medical technologies to be assessed, since OTA finds that this is the critical phase of any overall assessment strategy.

Identification

Any system for medical technology assessment must have mechanisms to identify technologies to be assessed and to set priorities among candidates for assessment. Clearly, no single mechanism is appropriate for all occasions and all technologies. What works for drugs may not be suitable for surgical procedures, and what is appropriate for identifying emerging technologies may not be adequate for established ones.

Methods of identifying technologies for assessment can be thought of as falling into one of three generic categories: 1) routine mechanisms, 2) priority-setting mechanisms, and 3) mechanisms of opportunity. *Routine mechanisms* systematically identify a class of technologies and are usually connected with a particular event with which all technologies in the class are associated. (Examples are FDA's requirement that all drugs and devices be registered with it prior to marketing or testing in human beings and, if taken advantage of, HCFA's reimbursement coverage determinations.) *Priority-setting mechanisms* are not routine and are often mechanisms or processes used by some group to determine priority technologies for assessment based on some implicit or explicit criteria. (Examples are the processes the institutes of NIH and HCFA use to establish their research agendas, the processes the Office for Medical Applications of Research (OMAR) of NIH uses to set priorities for consensus development conferences, and the process the National Center for Health Care Technology (NCHCT) formerly used to establish priorities for technology assessments.) *Mechanisms of opportunity* are means for identifying technologies for assessment as opportunities happen to occur. These are less well defined than mechanisms in the previous two categories, but are not necessarily less important, because technologies that suddenly become important to assess often do so for safety or ethical reasons. (Examples of mechanisms of opportunity are FDA's spontaneous drug reporting system, PSROs' medical care evaluation studies, and the systematic generation of patient outcome data which researchers can analyze or the public can challenge.)

The specific purpose of identifying technologies for assessment will vary according to the stage of development of the technology in question. Emerging technologies, especially drugs and certain devices, need to be assessed for safety and efficacy. In the case of drugs and certain devices, assessment in the emerging stage is necessary because of the well-accepted social responsibility, as expressed in FDA law, to protect the public. There may be other social and ethical reasons to assess emerging technologies as well. New and existing technologies need to be assessed for safety and effectiveness, sometimes for cost effectiveness, and at times for social and ethical consequences. Physicians and their patients need to know what works and what the benefit/risk balance is. Patients and insurers, including HCFA, need to understand the economic implications of technologies, especially when there are alternatives. Information is also needed by PSROs. PSROs are charged with assessing whether HCFA funds are being used on "needed" services, and these include both new and existing technologies. Additional information is needed by health planning agencies, which are charged with determining whether major new technologies should be purchased by hospitals.

A critique of the current system's record in identifying medical technologies for assessment yields mixed results. For the purposes of the discussion below, it is helpful to think of the identification of technologies at the four typical stages of development defined in chapter 1: 1) emerging technologies, 2) new technologies, 3) existing technologies, and 4) new applications of existing technologies.

Identification of Emerging Technologies

Certainly, the most thorough system for identifying emerging medical technologies is FDA's premarketing approval process for drugs and Class III medical devices (see ch. 6). This process is clearly a routine identification mechanism, as defined above.

The only other notable systems for identifying emerging medical technologies are priority-setting mechanisms. These include the processes for establishing the research agendas of NIH, the National Center for Health Services Research, and while it was funded, NCHCT. Although each institute and research agency has its own internal systems, the process of establishing priorities for intramural research and extramural contracts is essentially determined by institute or agency staff. Research priorities tend to be set by informed professional staff who know their particular field well and thus know which questions are important to address. Grants can be either solicited or unsolicited. In either case, the projects are generally selected on the basis of technical merit and judgments about their importance. The research agenda priority-setting processes of NIH and other Federal research agencies generate information for a base of knowledge which can lead to unpredictable but substantial future dividends that may be difficult or impossible to measure. Often, however, the processes do not address the immediate policy priorities of operating agencies such as HCFA and other social priorities such as Congress may have for the health care system.

Two other priority-setting mechanisms for identifying emerging technologies for assessment are also deserving of mention. HCFA's research arm, the Office of Research and Demonstrations, is charged with assessing technologies of interest to its operations. Some of these technologies may be classified as emerging, although most probably would not be. Seldom, however, are the technologies clinically related (e.g., some are concerned with information systems). The other mechanism was the NCHCT's "emerging technology list." This was a systematic and broad-based approach for identifying health-related technologies under development which were expected to be used in the practice of clinical medicine within 5 years. However, critics from industry charged that the compilation of such a list threatened the innovation process, and the 1981 reauthorization of NCHCT specifically withdrew the center's authority to compile this list.

The third type of identification mechanism, taking advantage of unforeseen "opportunities," is generally not relevant for emerging technologies, except in rare special cases such as the artificial heart.

Discussion: FDA's routine identification of all emerging drugs and emerging Class III devices seems adequate and appropriate. Emerging medical and surgical procedures do not seem to lend themselves to being identified through routine mechanisms. The most appropriate identification method for emerging procedures would seem to be the subjective priority-setting mechanisms such as those being used by the institutes of NIH and other Federal research agencies.

NCHCT's "emerging technology list" had the advantage of cutting across categorically related programs and also forced each program of the Department of Health and Human Services to explicitly identify technologies which were emerging and of importance. The 1980 format for submission of an entry included the name and identification of the technology, a technical description, a statement of importance or potential impact, an evaluation of the technology's present status and data needs, and any special considerations. NIH staff have commented that the exercise of compiling such a list was useful in taking stock of what was happening in their respective fields. If one objective of the technology assessment system is to more actively manage technologies, compiling such a list would seem to be quite useful in that it allows one to make predictions and to plan for the future. The charge of industry that the list inhibited innovation is not supported by any data that OTA could find, but the issue could be further examined.

OTA concludes that emerging drugs and devices are adequately and appropriately identified, but that emerging medical and surgical procedures could be better identified. Overall, the identification of emerging technologies is not a critical weakness of the present system.

Identification of New Technologies

As a group, new technologies, those in the adoption phase, are the most easily identified. In particular, such technologies are the most obvious candidates for identification through routine mechanisms. Most new medical devices (i.e., Class I and II devices) are routinely identified as required by FDA law (see ch. 6).* New medical and surgical procedures, including the use of new drugs and/or devices, are potentially identifiable routinely through the reimbursement process, since the question of coverage should arise when a new procedure is identified.

All new medical technologies are also potential candidates for being identified through the priority-setting mechanisms discussed in the preceding section on emerging technologies. And, in fact, new technologies are identified for assessment through the priority-setting processes of the institutes of NIH and the other research agencies of the Public Health Service (PHS) and HCFA.

All new technologies are also logical candidates for being identified through mechanisms of opportunity. The recent maternal serum alpha-fetoprotein (MSAFP) controversy illustrates the use of such a mechanism (see app. E). FDA was on the verge of approving widespread use of the MSAFP screening test when a special interest consumer group (the parents of children with spina bifida) questioned the validity of FDA's data. A major assessment of MSAFP was subsequently carried out, and new regulations were issued. This case illustrates a mechanism of opportunity (i.e., publishing data and making decisions under public scrutiny).

Discussion: As a class, new technologies are the most easily identified as candidates for assessment, especially by routine mechanisms. In the case of new Class I and Class II medical devices, FDA's routine identification process seems adequate and appropriate. In the case of new medical and surgical procedures, however, there is currently no systematic mechanism for identification. To some extent, new procedures are identified through the reimbursement system; however, in contrast to the structured identification process of FDA, identification through reimbursement decisions of HCFA and other public and private insurers is much more haphazard. While there is considerable potential for the reimbursement system to be used routinely as a primary means of identifying new procedures for assessment, problems persist. Even the process of identifying which procedures are new seems to be unsatisfactory: terminologies and codes on claims forms are often not accurately labeled or are not standardized; new procedures often do not have a procedure code number. However, to the extent that there is increased scrutiny by third-party payers of bills submitted for new procedures and more than occasional denial of payment for such bills, the provider has a strong incentive to request payment for an already existing standard procedure, rather than a new one, thus complicating the identification process.

---

*More invasive devices (i.e., Class III devices) are identified in the "emerging" phase, because they must receive FDA's approval before being tested in clinical trials.

The process of identifying new technologies through the priority-setting processes of the institutes of NIH and agencies of PHS has essentially the same strengths and weaknesses as were discussed in the connection with the identification of emerging technologies. From an academic point of view, the system seems appropriate. The weakness stems from the lack of an adequate system to identify priority candidates for the operating agencies, especially HCFA, PSROs, and planning agencies. Theoretically, HCFA has its own research arm, the Office of Research and Demonstrations, to accomplish this. As stated previously, however, that office has not been very involved to date with either identifying technologies for assessment or assessing them.

Whether the mechanisms of opportunity for identifying new technologies are adequate is difficult to assess. Since standardized, high-quality data on technology use and health outcome are not generally available, it is likely that they are not.

OTA concludes that new drugs and devices are adequately identified for assessment, but that new medical and surgical procedures are not. The most pressing need is for a routine mechanism to identify new procedures before they are widely adopted. The reimbursement system, because coverage and payment decisions are critical points in the diffusion of many technologies, might be given primary consideration. In addition, the priority-setting systems of the institutes of NIH and of other Federal research agencies (e.g., NCHSR) are adequate and appropriate for their respective mandates, but there is not an adequate similar system to fulfill the needs of operating agencies (e.g., HCFA, planning agencies). Finally, sufficient mechanisms of opportunity for identifying new technologies do not exist but could be developed. Medical specialty societies could be helpful in this area.

Identification of Existing Technologies

As a group, existing medical technologies tend to be the least likely candidates for routine identification, primarily because there is no natural triggering mechanism such as introduction. Consequently, the timely identification of existing technologies must depend largely on priority-setting procedures or mechanisms of opportunity.

Theoretically, if emerging and new technologies had been adequately identified (and assessed) as they developed, there would be less need to identify (and assess) them after their adoption and general diffusion. But, as indicated in previous OTA reports (e.g., 266,270,279), most existing medical technologies have not been adequately assessed. At a minimum, existing medical technologies should be monitored for risks that may not have been previously apparent. A review of the activity in this area reveals a very poor record, with a few exceptions and a few encouraging signs.

One encouraging sign is the interest in postmarketing surveillance systems for drugs.* Postmarketing surveillance systems are noteworthy, because there is increasing concern that FDA's premarket approval process is not sufficient to protect the public after a drug or device is marketed and in use (281). Although often regarded as testing techniques, postmarketing surveillance systems can also be thought of as sophisticated systems for identifying technologies needing further investigation. Such systems represent a hybridization of a routine mechanism, a priority-setting mechanism, and a mechanism of opportunity. Although data may be collected routinely under postmarketing surveillance systems, not all drugs would automatically be screened. FDA can set its own research agenda, and independent research investigators, at their own initiative, can be ex-

---

*This topic is considered at greater length in a separate OTA report entitled *Postmarketing Surveillance of Prescription Drugs* (281).

pected to use the data to identify fertile areas for future study.

Another encouraging sign for the identification of existing technologies for assessment are private sector initiatives using the priority-setting method. These include Blue Cross/Blue Shield's Medical Necessity Project and the American College of Physicians' new Clinical Efficacy Assessment Project (see ch. 6). In the Federal sector, discussed previously, priority-setting processes (including NCHCT's and OMAR's) are also used to identify existing technologies.

The most glaring omission in the system for identification of existing technologies for assessment is the lack of identification by operating agencies, especially HCFA and State Medicaid agencies. Even with its PSRO arm, HCFA does not have an adequate system to question technologies that are already in widespread use.

One mechanism of opportunity that can be used to trigger identification of an existing technology in need of assessment is the identification of a competing technology. To some degree, this mechanism is used implicitly. For instance, computed tomography scanning was likely to have been compared with existing technologies such as skull X-rays. Whether such opportunities for identification are always, or even frequently, taken advantage of is not clear.

It was stated earlier that mechanisms of opportunity are particularly useful for identifying technologies that are currently in use. FDA has a spontaneous reporting system for adverse drug reactions which illustrates how technology assessment opportunities surface "spontaneously." Similar systems could be used by other agencies such as HCFA. A functioning identification system of opportunity requires a method by which a technology assessment issue can be reported and a means to act on that information.

OTA concludes that the system for identifying existing technologies in need of assessment is inadequate. One promising possibility is postmarketing surveillance techniques. As was true with emerging and new technologies, the priority-setting procedures of Federal research agencies may be adequate for those agencies' respective needs, but not for the needs of operating agencies such as HCFA. And the operating agencies themselves do not adequately identify existing technologies for assessment. Medical specialty societies could be helpful in this area. Finally, NCHCT's activities of identifying nationally important priority technologies for assessment were valuable but are not now funded. Thus, no body is currently undertaking this important task.

Identification of New Applications of Existing Technologies

The consideration of new applications of existing technologies is important for two reasons. First, a new application of a technology means that previous information about the technology may no longer be applicable; and second, a technology's new use may provide an opportunity to identify it through a routine mechanism. At present, OTA is unaware of any systematic method of identifying new applications of existing technologies as candidates for assessment.

These technologies can be identified through priority-setting procedures and mechanisms of opportunity, as can existing technologies. It would seem, though, that the most rewarding approach for identifying new applications of existing technologies would be through a routine mechanism, namely, the reimbursement system.

OTA concludes that new applications of existing technologies are not adequately identified for assessment. To facilitate the identification of such technologies, the most promising approach may be the use of the reimbursement system to link the diagnosis with the use of the technology. Medical specialty societies could be helpful in this area.

Testing

Many of the deficiencies of the testing phase of the current system for medical technology assessment are intimately related to the inadequacies of the identification phase. In order to know what to test for, one must have identified the appropriate technology for assessment, the relevant policy concern (e.g., safety, efficacy, or cost effectiveness), and the information which is lacking. Thus, an adequate testing phase requires an adequate identification phase. It is not surprising, therefore, that the strengths and weaknesses of current testing activities closely parallel those of the identification phase.

FDA adequately identifies emerging and new drugs and devices that need assessment and also determines what information it needs. Furthermore, FDA carefully reviews the research protocols of the industries it regulates and requires that the protocols be used. Resulting testing by industry seems adequate. As suggested previously, however, FDA does not have an adequate means to identify which drugs and devices need further testing once they are released into the market. Thus, FDA cannot develop protocols for further testing of products in new settings or under different applications. As indicated in chapter 3, adequate protocols can be developed (see also ref. 281).

Chapter 6 discussed the testing of medical and surgical procedures through the funding activities of NIH. Since the individual institutes of NIH subject all research protocols to an intensive peer review process (270), the quality of the research is generally good. Any problems with such activities center around either the need for additional funding or the agenda-setting process* (the latter is essentially an issue of identification). It is important to note that NIH does not have the mission to ensure that all medical and surgical procedures are proven to be safe and effective. (Nor does any other agency or organization.)

Currrently, the overriding weakness of the testing phase is in the testing of new and existing medical and surgical procedures. Since procedures tend to be developed within the practice of medicine, they are generally adopted and accepted by the medical community without a routine, formal examination of their merits. A good deal of the problem in this area stems from a lack of research funding. Another problem concerns the development and use of research methods, since RCTs are not appropriate for all clinical inquiries.

Data systems can be linked and then used to identify technologies for assessment, and such systems can also be used, though to a lesser extent, to evaluate safety, effectiveness, and cost effectiveness. Prospective studies could be initiated to link technology use to health outcome and cost. One model which could be further examined is the Clinical Data Acquisition Plan which was being developed by NCHCT (see ch. 6). Data systems may be adequate in some cases to provide sufficient evidence of safety and effectiveness of technologies, especially if they are used to complement more rigorous testing methods.

FDA's postmarketing surveillance activities for drugs, mentioned earlier, are being developed to monitor adverse reactions to drugs (281). Such systems may be adaptable for other technologies as well.

The Federal Government has not used its potential leverage to test technologies through the reimbursement system. For instance, if HCFA could use its system to study whether new procedures were safe, effective, and cost effective, or needed further testing before final reimbursement decisions were made, many ineffective technologies might be identified and discarded well before they were accepted by the medical community. The ambiguous "reasonable and necessary" clause of HCFA's statutory language has been an obvious impediment to such activity.

Although the private sector has been actively involved in testing medical technologies, its direct support for well-controlled clinical trials has not been very extensive (except for the research monitored by FDA). Research protocols tend to be of a nonrandomized design and often rest on the information derived from available data bases and recordkeeping systems (e.g., 209). Nevertheless, there is evidence of increasing private sector interest in research on technologies. Much of the interest seems to stem from the belief and concern that resources are not used efficiently.

Finally, it should be noted that currently no public or private body has responsibility for determining either the cost effectiveness or social/ethical implications of medical technologies. FDA and NIH are both primarily oriented towards safety and efficacy issues. It is true that NCHSR and to a lesser extent HCFA do selectively fund some cost-effectiveness studies, but no one body is charged with systematically examining the larger social issues.

---

*During the current period of fiscal restraint, substantially increased Federal research budgets seem unlikely; this, it may be worthwhile to explore the possibility of joint private/public efforts. The theme is explored in ch. 8.

OTA concludes that, in general, drugs and devices are adequately tested for safety and efficacy prior to being marketed. Medical and surgical procedures, however, are not well tested for either safety or effectiveness. No class of technologies is adequately evaluated for either cost-effectiveness or social and ethical implications. Finally, there is no organization whose mission it is to ensure that medical and surgical procedures are assessed for safety and efficacy or to evaluate any class of technologies for cost effectiveness and for social/ethical implications.

Synthesis

Synthesis activities in the area of medical technology assessment are generally of two major types: 1) synthesis of the results of individual research studies; and 2) synthesis of a body of research findings with other concerns such as risk, social, ethical, or cost factors. The former, which is more focused and technical than the latter, seeks to answer questions such as those concerning the safety, efficacy, or effectiveness of a given technology. The second, which is more policy directed, often seeks to develop guidelines or standards for medical practice or reimbursement policy. The value of the latter depends, in large part, on the adequacy of the former. That is, one cannot consistently set good policy regarding medical technologies without knowing what the collective research says about a given set of issues.

The challenge for synthesizing research evidence concerning a technology is to make sense out of a growing body of information—some bad, some good. Techniques available to do this were described in chapter 5.

Synthesis activities are inherently a part of conferences sponsored by individual institutes of NIH and other Federal agencies (and numerous other organizations). Among the more formal synthesis-type activities within the Federal Government are the consensus development conferences sponsored by OMAR of NIH.

The goal of consensus development is to synthesize the scientific literature on safety and efficacy/effectiveness and to recommend to physicians the appropriate use of technologies. In many respects, consensus development conferences are well done and important activities. As discussed in chapter 5, however, the NIH consensus conferences have demonstrated weaknesses in terms of objectively synthesizing scientific information and in recommending guidelines for the appropriate use of the technologies they consider.

For instance, although the NIH panels are generally composed of eminent physicians, a methodologist (i.e., a biostatistician or an epidemiologist) is not always included, and the validity of evidence from scientific research is not always explicitly examined (see app. C). Thus, the methodological limitations of a given study may be overlooked. Another limitation of NIH's format is the process itself. For instance, the use of adversary groups and task forces has been almost entirely abandoned, and the questions that have been posed are strictly limited to issues on which there is enough factual evidence to reach agreement. For the purpose of synthesizing available knowledge, this approach may be adequate (assuming that the available knowledge is all included and understood), but for the purposes of identifying gaps in knowledge and needs for future research, this approach is weak. Of equal importance is that consensus development conferences tend to examine in depth only two aspects of medical technology assessment: safety and efficacy. This limits the usefulness of the conferences and calls into question the appropriateness of their setting guidelines for clinical use of a medical technology (e.g., frequency of Pap smears or the use of mammography).

Setting medical standards (e.g., indications for using respiratory therapy) by professional organizations and governmental agencies, though not customarily characterized as a synthesis activity, does depend on the integration of available information. Ideally, these organizations and the individuals within them should first systematically and objectively review the clinical research evidence. A knowledge base (see ch.5), such as the National Library of Medicine's (NLM's) Hepatitis

Knowledge Base, may be useful in this regard. An important output should be the identification of fertile areas for further research. However, the common pattern is for standards to be set, whether by PSROs, HCFA, professional organizations, or NIH, that are based on the group's belief of good medical practice, much of which is unsupported by scientific evidence. Thus, not only are important opportunities lost for further research, but perhaps more important, current practice patterns tend to be validated when they should not be. Finally, not only is the research evidence generally not reviewed systematically and objectively, neither is the standard-setting process. Formal decision-assisting techniques such as Delphi and nominal group techniques are seldom applied.

OTA concludes that the synthesis phase of the present system of technology assessment is unnecessarily weak within both the private and public sectors. Research evidence regarding the safety, efficacy, and effectiveness of medical technologies is seldom examined systematically and objectively. Federal agencies and private insurers and organizations set policies, guidelines, and regulations, and/or make reimbursement coverage determinations, many of which profoundly affect the adoption and level of use of medical technologies. Yet, their decisions are usually based on informal, subjective, group-generated norms which tend to support the status quo. Formal, more objective techniques both for evaluating research evidence and for making decisions and setting policy could be used more often to aid in better decisionmaking.

Dissemination

The issues associated with making sure that the right people have access to technology assessment information transcends technology class (i.e., drug, device, procedure). However, the dissemination issue is particularly important for the decisionmaker at the point of a technology's adoption. At that point, the insurers, hospitals, physicians, or patients need to assimilate safety, efficacy, and cost information in order to make a rational decision based on their individual conditions, values, and objectives.

This report does not deal with the entire scope of information transfer. It does, however, briefly examine the ability of the Federal Government to make available research findings and the activities of NLM in indexing and providing access to the biomedical and other health-related literature that may be useful for medical technology assessment. These issues are addressed in greater detail in a separate technical memorandum entitled *MEDLARS and Health Information Policy* (276). That document also discusses the relationship between NLM and private sector organizations that index and provide computerized access to the biomedical and other health-related literature.

Specific problems associated with communicating information about medical technologies appear to be similar to those in other fields of science and technology. Paradoxically, the amount of information available is at once too much and too little. The "publish or perish" syndrome has led to an explosion in the quantity of literature without an accompanying improvement in quality. One way to ameliorate the problem of an overabundance of primary literature has been to rely more on secondary sources, particularly bibliographic data bases that can be read by a computer.

NLM has excelled in collecting, indexing, and making accessible biomedical literature by a computerized bibliographic system. An earlier OTA staff paper of this assessment indicated that about 76 to 98 percent of the relevant biomedical journals are covered by NLM's major biomedical data base MEDLINE (278). But subject coverage of the health care delivery field by contrast is poor: less than 40 percent of all relevant citations were contained in MEDLINE. Many of the missing citations were in economic, business, and sociological journals. The coverage of citations in the health care delivery field is limited, not only because many of the citations are in economics and business journals, but because a large number are also older than the NLM health file. The percentage of relevant citations held in MEDLINE will be significantly greater for articles citing the more recent literature.

References to other sources of information on medical technology assessments such as monographs, reports, conferences, and Government documents are not nearly as accessible in other bibliographic data bases as references to the journal literature. Thus, many useful Government research documents may not be used or may have to await for the authors to publish the results in a refereed professional journal.

Along with the growth in literature in the biomedical field has come confusion on the part of many users about obtaining information. The large number of primary publications and even secondary publications (e.g., bibliographies) makes it difficult for the occasional user to find information efficiently. Users with access to a well-trained and competent information specialist or librarian find their search simplified. However, the quality of libraries or information centers and the quality of the staff vary. Furthermore, there is no comprehensive single source where information about existing federally generated data bases in a field can be obtained. This complicates even an informed user's search and has resulted in the unnecessary duplication of information.

Two important issues related to NLM's usefulness in the dissemination of technology assessment information are: 1) whether NLM should include more Government reports and other nonserial literature (especially in the area of health services) in its data bases, and 2) whether NLM should modify its indexing process to indicate more useful information as to articles dealing with research findings. With regard to the first issue, it should be noted that the National Technical Information Service (NTIS) has major responsibility for Government reports and perhaps NLM should not duplicate the collection, although NLM is expanding its data base somewhat in this direction. An effort could be made to link existing data bases so that a single search could access both NLM and NTIS data bases as well as any other sources relating to health questions. With regard to the second issue, one possibility would be for NLM to carry a code within its citations that is related to the methodological and statistical nature of the article. The editors of research journals could be asked to supply the necessary information (276).

Finally, the potential impact of the widespread distribution of microcomputers in physicians' offices in the future could be significant. For instance, NLM's data bases could be immediately accessible, and if knowledge bases such as NLM's Hepatitis Knowledge Base were available, the dissemination of technology assessment information could be much enhanced.

OTA concludes that better methods need to be found to communicate information about medical technologies to physicians, researchers, and policymakers. OTA also concludes that Government-generated reports, many of which may be important to technology assessment, are not as accessible as they could be. There is no mechanism through which all health-related Government reports can be identified or obtained. Finally, NLM's mission and capabilities should be examined to determine whether more Government reports should be included in its data base, and whether NLM should index articles to indicate their methodological and statistical nature.

CONCLUSION

Thus, OTA finds that there are major problems with each of the four components of the present system of medical technology assessment. The last chapter of this report provides Congress with options to address what appear to be some of the most striking weaknesses.

8.
Policy Options

> The great end of life is not knowledge but action.
>
> —*Thomas Henry Huxley*

Contents

Page
Introduction 105
Legislative Options 106
  Organizational Options 106
  Research Funding Options 107
  Educational Options 108
Congressional Oversight Options 108
  Private Sector Oversight Option 108
  Identification of Medical Technologies for Assessment 109
  Testing Medical Technologies 109
  Synthesizing Research Information and Group Decisionmaking Activities 109
  Dissemination Activities 109
  Establishing a Coordinated System of Technology Assessment 110

8.
Policy Options

INTRODUCTION

As described in the previous chapter, the present "system" of assessing medical technologies exhibits deficiencies in a number of areas. One of the problems is that there has been no strategy or systematic plan for developing an effective system:

1. to identify technologies to be assessed;
2. to ensure that high-quality, relevant assessments are carried out;
3. to synthesize or coordinate the synthesis of the resulting information; and
4. to disseminate the information to Federal agencies, health care providers, third-party payers, patients, and other health care decisionmakers.

Elements of an effective system are already in place—e.g., the Food and Drug Administration's (FDA's) processes for the regulation of drugs and the National Institutes of Health's (NIH's) support for clinical trials. The problem is that these elements have not become part of a coherent overall system. The most important need is to bring forth, from the present multiplicity of agencies and activities, a more rational and systematic approach to promote and coordinate medical technology assessment.

Achieving the goal of an effective system for assessing medical technologies will require a more integrated structure than now exists. An integrated system for assessing medical technologies need not be centrally managed or controlled. However, an integrated system will require strong links between multiple organizations and agencies. Candidate technologies for assessment could be identified by a number of Federal organizations—including NIH, FDA, Professional Standards Review Organizations (PSROs), health planning agencies, the Health Care Financing Administration (HCFA), the Veterans Administration, and the Department of Defense—as well as private sector organizations. To ensure that the most significant technologies are assessed, all involved organizations could participate in a priority-setting exercise. Many medical technologies needing assessment are already in widespread use. In setting assessment priorities, therefore, it might be useful to establish new links to nongovernment bodies such as medical specialty societies.

Mechanisms to fund assessments of high-priority technologies would have to be developed. Federal research organizations, such as NIH and the National Center for Health Services Research (NCHSR), should be involved. Private organizations may also be interested in participating in assessments.

An important function of any system for assessing medical technologies would be to select appropriate testing methods. Although the randomized clinical trial (RCT) is accepted as the optimal method for testing efficacy in most situations, resource and other constraints make it impossible to test every technology by this method. In some cases, alternative study designs may be more useful.

Syntheses of information could be done by both Government and private organizations to meet their respective needs. Information could be fed back to organizations participating in the assessment process in a form most useful or acceptable to them. HCFA, for example, in making reimbursement decisions, might be most interested in the question of whether, on the basis of scientific evidence, a specific technology could be considered to be efficacious.

To imagine how a coordinated technology assessment system could work, consider the assessment of a hypothetical high-priority medical technology about which relatively little is known. First, it would be necessary to gather and synthesize information about the technology. The process of synthesis, by pointing to gaps in available knowledge about the technology, might suggest a need for further research. It might be, for example, that the extent of use of the technology is not known; in that case, a simple data-gathering exercise by the health insurance system might be useful. It might be that the technology had been tested in normal subjects, but not in elderly people with chronic disease; in that case, its safety in the latter might need to be investigated by surveillance.

All agencies and organizations participating in the system could contribute to the assessment of this technology. Thus, for example, HCFA could provide information from its data base about the extent of the technology's use. If questions arose concerning benefits in the usual practice of medicine, selected PSROs might be asked to evaluate these. Different testing methods could be used simultaneously to complement one another. For instance, a small RCT could be used to establish causation, while an observational survey could be used to detect associations within a more diverse population. Unlike RCTs, which generate their own data for analysis, observational studies typically rely on existing, often large-scale, data collection systems (e.g., Medicare claims files, vital statistics). For purposes of analysis, it is important that these data systems be compatible with one another and be accurate.

When policy decisions about the technology needed to be made, the evidence of safety, efficacy, and effectiveness would be synthesized, and the information disseminated to the appropriate decisionmakers. The system would require mechanisms to determine when a rigorous assessment was needed and when a more informal review was sufficient. If controversy existed concerning appropriate patterns of use for the technology, group decision techniques such as those described in chapter 5 would be useful.

The initial concept of the 1978 legislation establishing the National Center for Health Care Technology (NCHCT) resulted from a recognition in Congress of the need for a systematic approach to the assessment of medical technologies. However, the NCHCT legislation left certain problems unaddressed, e.g., who would set research priorities for the Government. Furthermore, NCHCT's mandate to perform assessments was curtailed by its austere budget. Consequently, NCHCT's impact on the health care system has been fairly small. If NCHCT's funding is not restored, however, an organization potentially able to carry out or coordinate the tasks mentioned above will have been lost.

The policy options that follow are intended to address the deficiencies of the existing system for assessing medical technologies. The options are divided into two broad categories: legislative and oversight. OTA finds that there are few realistic legislative options necessary for Congress to consider. In most of the deficient areas noted within this report, congressional oversight may suffice. There is already substantial statutory authority vested in the Secretary of Health and Human Services to develop a coherent system of medical technology assessment. The options below are not presented in any particular order of importance, nor should they be regarded as mutually exclusive.

LEGISLATIVE OPTIONS

Organizational Options

1. **Sponsor a private-public body or grant a charter to an organization to undertake medical technology assessment activities.**

An organization could be chartered either as a separate nonprofit corporation or as part of an organization (e.g., the Institute of Medicine of the National Academy of Sciences) to undertake assessment activities that would complement Federal activities and serve the needs of consumers, providers, and third-party payers. The organization could be composed of a number of groups concerned with the evaluation of health care: physician and hospital professional associations, consumers represented through industry and labor, private health insurers, and academic centers.

One of several objectives that such an organization could have would be to stimulate the development of uniform and accessible data bases for medical technology assessment. This could include encouraging the use of uniform diagnostic and procedure coding, encouraging commonality of patient registration and claims forms, and developing clinical data banks. A second objective would be to identify technologies for assessment and to establish assessment priorities. A third would be to develop and refine methods of assessment, including scientific, economic, and social tools. This objective could include development of community-based collaborative studies, improved clinical data banks, better measures of quality of life, etc. A fourth objective would be to conduct comprehensive assessments of medical technologies, considering their scientific, economic, social, ethical, and legal implications; and to perform scientific and economic analyses at the request of providers and third parties. The performance of such assessments would include the generation of new data as needed. A fifth objective would be to disseminate new information and to serve as a clearinghouse of information on new technologies, assessments of technologies, etc.

Initial funding for the organization could come from private foundations. Ongoing support might include some support from foundations, contributions from insurers for support of assessment activities, congressional appropriations for special assessments of interest to the Federal Government, and support from hospital associations for advice on use and distribution of technologies.

One of the advantages of this option's general approach is that it would capitalize on private sector initiative and interest and would rely on private as well as possible public funding. A combination of private and public sector involvement may be essential for any system of medical technology assessment to be acceptable to all parties concerned. Apart from the very real possibility that an effective arrangement could not be forged, disadvantages of this approach include potential legal problems with funding—e.g., possible, though not likely, antitrust violations, and interference with State laws governing the health insurance industry.

A variation of this option, presented in appendix F, would establish a private-public body termed an "Institute for Health Care Evaluation." A limitation to the particular model proposed in appendix F is that it deals with medical technology assessment primarily as it relates to the reimbursement system. Thus, it may be unnecessarily restrictive. Both the legal issues noted above and ethical concerns associated with selectively reimbursing for health care technologies are discussed in appendix F.

2. Maintain the authority of and fund NCHCT.

Several advantages would result from refunding NCHCT. In the few years of its operation, NCHCT was making progress on several fronts. Perhaps most importantly, the Technology Coordinating Committee of the Department of Health and Human Services (DHHS), chaired by the Director of NCHCT, provided a valuable framework for the coordination of technology assessment within the Government; NCHCT's conference (e.g., on coronary artery bypass surgery) were successful as a needed adjunct to the more medically oriented NIH conferences; NCHCT provided an important focal point for HCFA to interact with the Public Health Service (PHS) for coverage determinations. Refunding NCHCT would allow it to continue this work and to mature as a Federal agency. Furthermore, even if option 1 above were implemented, the Federal involvement would still require interagency coordination.

The disadvantages of this option include most of the arguments which recently led Congress not to fund NCHCT for fiscal year 1982. A major concern at that time were the assertions by the medical devices industry that NCHCT's "emerging technology list" inhibited innovation. The other major concern was that NCHCT's activities might not be needed, because professional medical societies are increasingly active in technology assessment and PHS may be able to manage many of NCHCT's former responsibilities.

Research Funding Options

3. Change the statutes so that HCFA can selectively reimburse for experimental technologies in return for clinical data on these technologies.

This option has several potential advantages. First, the actual implementation of this option would not necessarily involve additional costs. Second, the implementation of this option might prove over the long run to be an effective method of cutting costs. Decisions to reimburse for many technologies which are essentially experimental are now made before adequate safety, efficacy, and cost-effectiveness information is available. If implemented properly, this option could substantially increase the quality of information available for reimbursement coverage decisions, thereby yielding substantial budgetary savings.

The possible disadvantages of this option are also substantial and are similar to some of those of option 1. Primarily, the problems concern the legal and ethical implications of selectively reimbursing for health care. Before Congress seriously considers exercising this option, therefore, it would probably need to conduct extensive hearings concerning possible adverse consequences. Possibly, elements of PHS could be involved in developing research protocols and in interpreting research evidence from the resulting experiments. If option 2 above is exercised, NCHCT could perform these duties.

Educational Options

4. **Increase funding to train researchers in methodological and statistical principles.**

This option is a general one that could be accomplished through a variety of existing educational programs. One advantage of this option is that the quality of both privately and publicly funded research could be expected to improve over time; the quality of the synthesis of research findings could be expected to improve as well. The disadvantages are that this option would require additional funding and would not produce immediate results.

5. **Increase efforts to train health professionals in methodological and statistical principles.**

This option could be exercised either by categorical funding for additional training or through congressional oversight with respect to the educational curricula of professional and continuing educational programs. One advantage of this option is that it would help to increase the quality of research performed by clinical professionals. Perhaps more importantly, it would help to ensure that such professionals are more informed about the value and limitations of research literature in their respective fields. Disadvantages might be the cost of increased training efforts and the lack of immediately observable results.

CONGRESSIONAL OVERSIGHT OPTIONS

An option involving the private sector and eight other options involving the powers already vested in the Secretary of the Health and Human Services are discussed below. Congress could exercise these options by using its oversight powers.

Private Sector Oversight Option

6. **Encourage the private sector to take the lead in assessing medical technologies.**

As noted in chapter 6, there is evidence that the private sector is increasing its technology assessment activities. The advantages of this option are that it would require no additional funding, would probably be more attractive to elements of the private sector than other options, and would capitalize on an existing trend at an early stage. Disadvantages of this approach include the problem of differing private and public objectives; because of these differing objectives, much of the research conducted by the private sector may not be of high priority to Congress or the Secretary of Health and Human Services. A further problem may be a low level of funding. Since technology assessment is apt to be very expensive and since the information it produces is generally regarded as a public good, any *one* private party has an incentive to let someone else pay for it. Finally, the private sector does not have an impressive record in the assessment field; most past efforts have been Federal ones or have been required by Federal law.

Identification of Medical Technologies for Assessment

7. **Examine how Federal research institutes (e.g., NIH), agencies (e.g., NCHSR), and research programs of operating agencies within DHHS (e.g., the Office of Research and Demonstrations of HCFA) could identify technologies better when setting their research agendas; and how the PSRO program and the reimbursement system itself could be used to more advantage for identifying candidate technologies.**

As discussed in chapter 7, Federal programs do an inadequate job of identifying technologies which need assessment, especially medical and surgical procedures. This option is intended to address that problem.

Testing Medical Technologies

8. **Continue to conduct oversight hearings concerning the duplication and fragmentation of health-related data collection activities.**
9. **Examine the ability of operating agencies within DHHS (e.g., HCFA) to generate sufficient information for their decisions related to medical technologies, and the extent to which the Secretary of Health and Human Services utilizes the department's other research arms (e.g., NCHSR, NIH) to procure that information in a timely manner.**
10. **Examine the activities, plans, and potential for elements of DHHS (e.g., NIH) in utilizing various research methods to determine the appropriate use of medical technologies.**

The duplication and fragmentation of health-related data collection has been discussed in previous OTA reports and is well known to Congress. If Congress believes there is a continuing need to evaluate data collection activities and to match their value with both research and operating needs, it may wish to exercise option 8.

As discussed in chapters 6 and 7, operating agencies of DHHS (including HCFA), PSROs (within HCFA), and State and local health planning agencies require information for the decisions which they make, yet they all have limited ability (funding, expertise, and/or legal mandate) to secure such information. Option 9 would allow Congress to investigate this matter further.

Option 10 addresses research methods. As discussed in chapter 3, the selection of optimal research methods for evaluating different technologies at different stages in their lifecycle is very complex. Although an RCT is often ideal for studying efficacy, other methods may be more appropriate for such things as safety or for technologies in widespread use. This option would permit Congress to encourage the use of appropriate methods.

Synthesizing Research Information and Group Decisionmaking Activities

11. **Explore how research evidence could be better evaluated by HCFA and its carriers and fiscal intermediaries when making reimbursement decisions, by PHS when making recommendations to HCFA on coverage policy, by PSROs when setting standards for care, and by the Office of Medical Applications of Research of NIH when conducting consensus development conferences; and monitor the progress and potential costs and benefits of the National Library of Medicine's (NLM's) knowledge base prototype.**

As discussed in chapters 6 and 7, many decisions are currently being made by Federal agencies regarding premarket approval, reimbursement, and appropriate use of medical technologies. As discussed in chapter 5, however, there is good reason to believe that the evidence from research is seldom carefully and objectively analyzed before these decisions are made. This option would help ensure that these decisions are better informed and would assist in establishing research agendas for Federal agencies.

Dissemination Activities

12. **Examine the disposition of federally generated reports to determine the degree to which they have been useful both to private and public researchers and policymakers; specifically conduct an oversight hearing on the**

ability of researchers and policymakers to locate, retrieve, and use these reports.

13. **Examine whether NLM's literature base should be further expanded, especially to include more Government research reports and other nonserial literature; and examine whether there are more useful ways to index articles which contain findings from research.**

Federal agencies that conduct or fund research generate reports that may be useful for technology assessment. As discussed in chapter 2, numerous agencies and other organizational units of DHHS are involved in disseminating Government reports and other health-related information that is useful for medical technology assessment. The deposition of all research reports and other Government documents to distributing organizational units is not mandatory. Option 12 would allow Congress to ascertain whether, and the degree to which, federally generated information is useful and accessible to the people and agencies conducting medical technology assessments and making related health policy decisions.

As discussed in chapter 2, NLM is primarily oriented toward the biomedical research community. Increasingly, however, the health services research community is looking to NLM for assistance in locating and retrieving health services information. Option 13 suggests two areas that Congress may wish to explore.*

Establishing a Coordinated System of Technology Assessment

14. **Encourage use of the powers vested in the Secretary of Health and Human Services to develop a coherent system of medical technology assessment.**

As already discussed, a recent decision was made by Congress not to fund NCHCT. If Congress does not choose to restore NCHCT funding (see option 2), it may wish to consider this option.

---

*NLM's role in the dissemination of health-related information is explored at greater length in OTA's technical memorandum entitled *MEDLARS and Health Information Policy* (276).

Appendixes

Appendix A.—Compendium of Statistical Data Sources for Medical Technology Assessment

System/organization	Purpose and focus of data	Method of data collection	Population/location/time coverage	Statistics generated
Basic Vital Statistics NCHS	To provide uniform data on births, infant deaths, fetal deaths, birth weights, gestation.	Birth and death registration in States; 100 percent registration of all births and deaths in the country.	Total United States since 1933, registration for all locations on an ongoing basis.	Frequency of deaths by cause and fetal death. Frequency of birth.
National Health Interview Survey NCHS	To provide data on acute and chronic illness prevalence among noninstitutionalized persons.	Morbidity Survey—Multistage probability sample of standard geographic primary sampling units, census enumeration districts, and households. Interviews conducted in households.	120,000 civilian noninstitutionalized individuals throughout the United States covered each year since 1957. Data collected weekly throughout the year.	Estimated frequency of physician visits and hospitalization. Prevalence of chronic conditions. Incidence of acute condition. Morbidity associated with acute conditions.
National Health and Nutrition Examination Survey (NHANES) NCHS	To provide data on health status and physiological measures of noninstitutionalized persons.	Morbidity Health Survey—Multistage, highly clustered probability sample of persons stratified by geographic region and population density grouping. Primary sampling unit is the same as for National Health Interview Survey. Data collected by interview, physician examination, measurement and laboratory testing.	Two cycles of NHANES have been conducted on the noninstitutionalized U.S. civilian population. Cycle I 1971-74, Cycle I augmentation 1974-75, Cycle II 1976-79. A special survey of Hispanics is scheduled to begin in 1982.	Morbidity measures. Distribution of physiologic variables. Prevalence of chronic conditions.
National Hospital Discharge Survey NCHS	To provide medical diagnosis and surgical data on patients discharged from non-Federal short-stay hospitals.	Hospital Utilization and Morbidity Survey—Two-stage cluster probability sample. Hospital selection stratified by bed size, region, and ownership. Systematic sample of discharged patients.	Since 1964, over 400 hospitals surveyed per year throughout the United States.	Estimated frequency of specific diagnoses and surgical procedures by patient and hospital characteristics.
National Death Index NCHS	To provide possible fact of death, the death certificate number, and State of death.	Registry of all death records in the United States transmitted by the States or other death registration areas.	Information for all States beginning in 1979. Updated annually thereafter.	No statistics generated. Computer searches used to assist researchers to determine whether persons in their studies may have died.
National Natality Survey NCHS	To provide in-depth data on newborns and maternal health. Prenatal and postnatal care, infant health, medical aspects of pregnancy, labor, and delivery.	Followback survey. Probability sample of 1 in 425 live births drawn from birth certificates.	54 birth registration areas span the United States. Natality surveys done on records of 1963-69 and 1972. 1980 survey is currently underway.	Demographic statistics of parents. Frequency of radiologic procedures (1963). Factors associated with maternal and infant health.
National Fetal Mortality Survey NCHS	To provide data on the health of women who have stillbirths. Prenatal and postnatal care, medical aspects of pregnancy, labor, and delivery.	Followback survey. Probability sample of two in five fetal death certificates.	1980 survey, covering 52 birth registration areas, is currently underway.	Parental occupation. Fact of maternal radiologic exposure. Recent pesticides and insecticide exposure.
National Ambulatory Medical Care Survey (NAMCS) NCHS	To provide data on use of office-based physicians.	Morbidity Survey—Multistage probability sample of non-Federal physicians in office-based practice. A systematic sample of visits during a year within physician offices.	Since 1973, an annual cycle of NAMCS has been conducted on noninstitutionalized visits to non-Federal physicians in the 48 contiguous United States.	Frequency of visits by diagnosis. Visit rates to physicians by age, race, sex, and type of physician.
National Disease Surveillance Program CDC	To examine disease and trends of 45 specific conditions, to identify regional problems, and to evaluate the effectiveness of control measures.	Reporting System—Through epidemiology and laboratory offices of State health departments. Supplemented by information from epidemic investigations or outside sources such as NCHS.	State, county, and city health authority areas throughout the United States.	Incidence of 45 specific disease conditions.

System/organization	Purpose and focus of data	Method of data collection	Population/location/time coverage	Statistics generated
Birth Defects Monitoring System CDC	To provide data on type and incidence of birth defects.	Reporting System—Self-selected sample of hospitals. Commission on Professional Hospital Activities (CPHA) hospitals with obstetric units are requested to participate. Report aggregate data on birth defects.	Includes approximately one-third of all U.S. births. Quarterly reporting began in 1970.	Incidence of birth defects.
Hepatic Angiosarcoma Surveillance Finding Effort CDC	To understand relationship between angiosarcoma with health history and risk factors.	National case finding effort.	167 cases found, between 1964 and 1974.	Incidence, risk factors.
Smokey Nuclear Test Cohort Study CDC	To understand relationship between cancer, particularly leukemia, and nuclear test exposure.	Complete followup of all film badge holders.	Department of Defense workers at Smokey Nuclear Test Site in 1957 who worked 1 week prior to or after nuclear explosion and has possible exposure.	Incidence and mortality data.
Alcoholism Program Monitoring System NIAAA (ADAMHA)	To provide data to plan, manage and evaluate alcoholism.	N/A	Programs funded by NIAAA.	Data on clients, services, and costs.
Client-Oriented Data Aquisition Process (CODAP) NIDA (ADAMHA)	To provide data on federally supported drug abuse and rehabilitation programs.	Required reporting system.	Collection at time of admission to, and discharge from, treatment programs.	Client-related data.
National Drug Abuse Treatment Utilization Survey NIDA (ADAMHA)	To provide data on nationwide resources devoted to drug abuse treatment, their use, and their distribution.	Systematic survey, participation voluntary.	Annually—nationwide.	Use and distribution of resources for drug abuse treatment.
Drug Abuse Warning Network (DAWN) NIDA (ADAMHA)/DEA	To provide data on drug-related deaths, medical emergencies, and psychological crises.	Voluntary reporting system.	Participating emergency rooms, medical coroners, crisis intervention centers.	Mortality and morbidity incidence.
Adverse Drug Reaction Spontaneous Reporting System FDA	To provide for reporting of adverse effects of pharmaceutical products. Fact of death, health outcome after reaction, concomitant diseases.	Reporting system—Drug manufacturers, hospitals, health care practitioners file reports.	Since 1968, 167,000 reports have been filed across the United States.	Types of reactions by drug.
Poisoning Control Case Registry FDA	To provide registration of acute poisioning incidents. Signs and symptoms of poisioning; required medical intervention.	Registry—Voluntary reporting of all contracts with 400 poison control centers.	Since 1971, 1.2 million records of incidents have been placed in this system from across the United States.	FDA use only.
Registry of Tissue Reactions to Drugs FDA/Armed Forces Institution of Pathology under auspices of Universities Associated for Research in Education and Pathology, Inc.	To provide historical, clinical, laboratory, and morphological findings of adverse drug reactions.	Registry—Medical facilities reporting system.	3,753 cases from 37 States since 1966.	Quarterly reports of individual cases as to cause, basic disease, part of the body affected, and suspected drug(s).
Registry of Dermatological Reactions to Drugs FDA/National Academy of Dermatology	To improve the reporting of drug-caused skin reactions, and to test the feasibility of using a specialty society in medicine as the focal point for collecting drug experience data.	Registry—24-hour toll-free telephone reporting system.	National coverage; 3-year study starting in December 1980.	Incidence data.

System/organization	Purpose and focus of data	Method of data collection	Population/location/time coverage	Statistics generated
National Registry for Drug-Induced Ocular Side Effects				
FDA	To provide better information on ocular side effects of drugs.	Registry—Cases are reported primarily by opthamologists.	National—700 cases reported between 1977 and April 1978.	Quarterly and annual reports on individual drugs and effects.
Registry of Adverse Reactions to Contrast Media				
FDA/International Society of Radiology	To provide greater detail to the current study of adverse reactions to intravascular contrast media.	Prospective registry.	30 teaching hospitals in the United States, Canada, and Europe. 5,546 patients recorded with reactions out of 112,003 cases, time period not reported.	Incidence rates by types of examination and by types of patient reaction.
Nationwide Evaluation of X-Ray Trends (NEXT)				
Bureau of Radiological Health (FDA)	To detect and evaluate the extent of the population's exposure to X-rays used in medical and dental examinations.	Systematic sample for participating States.	Hospitals, private offices, and clinics.	Incidence rates.
Medically Oriented Data System (MODS)				
Bureau of Medical Devices (FDA)	To provide an estimate of the frequency of use of various drugs, devices, and procedures as well as the frequency of adverse events following from such use.	Systematic sample from participating hospitals.	Since September 1974, since 1977 has been a sample of 26 U.S. short-term hospitals.	Incidence rate.
Registry for Implanted Artificial Cardiac Pacemakers				
FDA	To provide data on new implant cardiac pacemakers, replacements, and intercurrent procedures.	Registry of three medical center experiences.	Medical Centers at University of Southern California, Montefiore Hospital, and Newark Beth Israel from July 1, 1974 to Dec. 31, 1979; 5,070 registered pacemakers overall.	Use patterns. Morbidity and mortality rates.
Surveillance, Epidemiology, and End Results Program (SEER)				
NCI (NIH)	To measure type and site of cancer incidence and survival in the United States; extent of disease; annual vital status.	Morbidity survey—Sample of persons diagnosed with cancer. 10 SEER locations chosen for contractual reasons; all cases of cancer in SEER area sampled.	5 entire States and 5 metropolitan areas have been surveyed on an ongoing basis since 1973. Some of these locations surveyed in earlier NCI cancer surveys. Population in these locations constitute 10 percent of U.S. total.	Survival rates. Incidence of cancer by site of cancer, geographic area, race, and sex.
Framingham Heart Study				
NHLBI (NIH)	To understand relationship between personal characteristics, physiologic measures, physical signs and heart disease, cause of death.	Prospective study—Representative sample of 1940 Framingham population. Followup of offspring of initial sample.	Followup of 5,209 persons in initial cohort ages, 29 to 62, in 1950, and 5,135 offspring, ages 16 to 52, in 1973. Followup began in 1940's and has continued up to present in Framingham, Mass.	Cardiovascular and cancer incidence. Risk statistics. Mortality rates.
Multirisk Factor Intervention Trial				
NHLBI (NIH)	To understand relationship between risk factors and heart disease; physiologic measures, physical signs and fact of death from heart disease.	Intervention trial—Two-stage sampling scheme. Nonprobability sample at 20 competitive research sites. Identification of males at high risk of heart disease.	Followup of 12,866 men, ages 35 to 57, from 1973 to 1976.	Not yet available.
Hypertension Detection Second Followup Program				
NHLBI (NIH)	Physiologic measures, physical signs, fact of cardiovascular disease and death recorded to understand relationship between drugs and control of hypertension.	Intervention trial. Two-stage sampling scheme. Nonprobability sample of 14 communities. Identification of persons with elevated blood pressure in a set time period.	Since 1973, followup of 211,000 persons, ages 30 to 69.	Morbidity and mortality rates.

System/organization	Purpose and focus of data	Method of data collection	Population/location/time coverage	Statistics generated
Lipids Research Clinics Trial Study NHLBI (NIH)	Physiologic measures, physical signs and cause of death recorded to understand relationship between drugs and control of hyperlipidemia.	Intervention trial. Two-stage sampling scheme. Nonprobability sample of 12 lipid research clinics. Identification of persons known to be hyperlipidemic.	Since 1973, followup of 214,000 men, ages 35 to 69.	Not yet available.
Coronary Artery Surgical Study (CASS) NHLBI (NIH)	To provide data on patients undergoing coronary arteriography from which pool of available patients for a randomized clinical trial on coronary artery surgery could be drawn.	Registry.	Initiated in 1973; 24,188 patients entered into registry, from this 780 patients recruited for clinical trial.	Incidence rates.
Percutaneous Transluminal Coronary Angioplasty (PTCA) Registry NHLBI (NIH)	To collect and report incidence data and protocols of a new medical device/procedure.	Voluntary registry.	International and ongoing; about 800 of 1200 procedures performed were reported by end of 1980.	Use patterns. Mortality and morbidity rates.
Multicenter databank networks for neurological disorders NINCDS (NIH)	To determine whether data collected routinely within the patient care process can benefit both clinical research and data management for 1) stroke patients and 2) traumatic coma victims.	Each involves a prospective databank network at four Federally-funded university based clinical centers in United States.	Implementation began in 1979.	Relationships between symptoms, stages of diseases, intervention, and outcome.
National Diabetes Data Group NIAMDD (NIH)	To estimate prevalance of diabetes on national scale, and monitor changes.	Statistical monitoring effort utilizing data pulled from existing files at NCHS.	National.	Incidence patterns.
Utah Geneological Data Base NIGMS (NIH)	To estimate and predict medical and genetic patterns, and to analyze existing geneological patterns.	Geneology was constructed from family group sheets reconstituted by members of the Church of Jesus Christ of Latter-Day Saints.	170,000 families in Utah.	Fertility, mortality, and other demographic data.
Medicare Claims File HCFA	To provide data on reimbursable claims from Medicare recipients; medical or surgical diagnosis; fact of death.	Registry—Universal sample of claims for short-stay hospitalizations, nursing home claims, and physician bills are gathered. A 20-percent systematic sample of claims are coded for medical diagnosis.	Since 1966, 20 million claims on people age 65 and over, those with chronic renal diseases, and those who meet the disability provisions of the Social Security Act. Data entered on an ongoing basis.	Utilization statistics by diagnosis.
Medicaid Management Information System (MMIS) HCFA	To provide data on reimbursable claims from Medicaid recipients; medical and surgical diagnosis; fact of death.	State-based data collection systems. 27 States have certified MMIS programs, 19 others are planning or implementing an MMIS.	Since 1966, over 27 million claims for medical care of low-income people.	Utilization statistics.
Professional Standards Review Organization (PSRO) Hospital Discharge Data Set HCFA	To provide data on patients reviewed by the PSRO review system; medical, surgical diagnosis and fact of death.	Registry—Universal coverage of patients reviewed by PSRO concurrent review process until 1977. A 20-percent systematic sample since 1977 using uniform hospital discharge survey forms.	Since 1975, PSRO's have provided data on Medicare and Medicaid discharges from 60 medical care regions.	Utilization statistics by diagnosis.

System/organization	Purpose and focus of data	Method of data collection	Population/location/time coverage	Statistics generated
Continuous Disability History Survey				
SSA	To provide data on applicants for disability benefits under Social Security Act, title II. Disability by diagnosis for individuals.	Survey of applicants for title II benefits. Probability sample of claims stratified by State.	Since 1967, 15,000 records have been placed in this file per year from throughout the United States.	Frequency distribution of disability-related primary diagnosis by State.
Leed File				
SSA	To provide integrated occupational and health status information on social security number holders. Fact of disability and death.	1-percent digital probability sample of social security numbers. Integration of work history from Annual Employee-Employer File with Summary Earning File. File provides longitudinal information by person.	Since 1957, 1-percent sample of wage and salary workers covered under Social Security.	Employment and earnings information from longitudinal analysis.
Annual Disability Determination/Social Security Income Extract				
SSA	To provide data on applicants for benefits under Social Security Act, title XVI. Disability by cause.	Survey of claimants for title XVI benefits. 10-percent probability sample of claimants stratified by State with oversampling of claims in small States or for children.	Since 1975, 100,000 records have been placed in this file per year from throughout the United States.	Frequency of disease and accident-related disability claims by State.
Census of the Population				
Bureau of the Census (Commerce)	To provide demographic data on entire population.	Enumeration of all residents.	United States, every 10 years.	Demographic and socioeconomic characteristics.
Annual Occupational Injuries and Illness Survey				
BLS	To provide data on work-related disease and injury. Acute disease and acute injury by cause, with or without lost work days. Fact of death.	Morbidity Record Survey—Probability sample of employers under OSHA recordkeeping, stratified by industry and establishment size.	Since 1972, ongoing surveys of nearly all private sector industries across the United States.	Incidence of illness and injury by type of case per plant-hour worked.
CHAMPUS (Civilian Health and Medical Program of the Uniformed Services) Management Information System				
DOD	To provide information for eligible beneficiary programs. Provider, claims, utilization and management data.	Registry of program recipients.	Dependents of active duty personnel, retired members of the armed forces, and others.	Utilization statistics.
TRI-Service Medical Information Systems (TRIMIS) Program				
DOD	Clinical and administrative automated data processing for military medical treatment facilities.	Patient registry.	80 military medical facilities; potentially about 8.9 million beneficiaries in DOD community.	TRIMIS systems in operation include Pharmacy Formulary, Hypertension Management, Hospital Logistics, and Automated Cardiac Catherization Laboratory.
Compensation and Pension System				
VA	To provide data on recipients of VA benefits. Disability by cause. Fact of death in or out of service.	Registry—Universal coverage of all veterans discharged with disability and receiving benefits.	Since 1960, all veterans discharged who receive benefits for military-related disability.	Level of compensation by diagnostic codes.
Patient Treatment File				
VA	To provide data on VA system discharges. Medical or surgical diagnoses.	Registry—Universal sample. Medical records abstracts for each hospital discharge.	Since 1969, all VA systems patients discharged.	Frequency of diagnostic category.
Blue Cross-Blue Shield Systems				
Over 100 members of Blue Cross-Blue Shield Association	To provide information on beneficiaries for reimbursement and utilization review/quality control generally.	Registry of claims forms for reimbursement.	More than 80 million people in the United States. Six million of these are Federal workers (which for some research, approximates a national data set).	Utilization statistics (data elements may vary from member to member).

System/organization	Purpose and focus of data	Method of data collection	Population/location/time coverage	Statistics generated
Computer-Stored Ambulatory Record System (COSTAR)				
Massachusetts General Hospital/NCHSR	To provide automated medical records and business support; quality assurance; patient followup reminders and selection of preferred therapies.	Data bank of beneficiaries.	Harvard Community Health Plan.	Utilization statistics.
Problem-Oriented Medical Information Systems (PROMIS)				
University of Vermont/NCHSR	To restructure medical records and data to organize and help direct the process of clinical care and medical action.	Computerized data bank of patient records entered by professional personnel.	Available since 1977.	Patient clinical and utilization statistics.
Hospital Discharge Data Systems				
An estimated 18 to 20 private nonprofit hospital discharge abstract systems, including the Commission on Professional and Hospital Activities (CPHA), and the Hospital Utilization Project (HUP)	To provide summary information about patients and their episodes of illness in short-term hospitals.	Registry—Patient information for participating hospitals is abstracted from medical records by hospital personnel after patients are discharged, according to a prescribed format.	About half the hospitals in the United States, representing about 20 million discharges annually.	Utilization and clinical statistics.
DES Vaginal Cancer Registry				
Dr. Arthur Herbst/ University of Chicago	To investigate the clinical, pathologic, and epidemiologic aspects of clear cell adenocarcinomas occuring in the vagina and cervix of females born after 1940.	Voluntary registry.	International in scope, with 341 cases reported; established since 1971. Considered highly precise, i.e., has captured about 85 to 90 percent of DES universe.	Mortality and morbidity data; clinical histories.
Rhode Island Diabetes Registry				
Rhode Island Deparment of Health	To provide information on young insulin-dependent diabetics (Type I juvenile diabetic).	Prospective 3-year registry. Charts reviewed on weekly or biweekly basis, with individual physicians subsequently interviewed.	Patients who are 1) insulin-dependent at 15 acute-care hospitals in the State, 2) under 30, and 3) residents of the State.	Incidence, patient-care patterns.
Pittsburgh Diabetes Registry				
N/A	To provide information on the etiology of juvenile onset diabetes.	Retrospective registry.	Pittsburgh, Pa., metropolitan area covering a 12-year period.	Incidence data.
American Rheumatism Association Medical Information System (ARAMIS)				
Consortium of several medical groups centered at Stanford University Medical Center	To collect patient information on a variety of rheumatic diseases among different patient populations.	Six data banks at medical institutions.	10,000 patients in system in the United States and Canada; extensive followup in some cases.	Incidence rates, clinical patterns.
Duke University Cardiovascular Data bank				
Duke University Medical Center/NCHSR	To provide information on patients with known or suspected ischemic heart disease and to describe outcomes of patients with various sets of attributes, patients findings, histories, outcomes.	Computerized data bank.	Over 6,000 patients at Duke University Medical Center since 1976.	Incidence and clinical patterns.
Databank on patients in coma for causes other than head injury				
New York Hospital-Cornell University Medical Center	To provide information on nontraumatic coma patients; allow future predictive power and clinical improvement in treatment; symptoms, diagnoses, procedures, functional states, causes of death.	Computerized data banks located in several medical centers in the United States and Great Britain.	Accesses about 500 patients per year.	Morbidity and mortality data; patterns of clinical practice.

System/organization	Purpose and focus of data	Method of data collection	Population/location/time coverage	Statistics generated
Intensive Care Databank				
Massachusetts General Hospital	To collect information and evaluate practices regarding intensive care units; diagnoses; functional status, indication data, charges.	Computerized data bank. Patient records and subsequent interviews were recorded and coded.	2,305 patients admitted to one of the three intensive care units at Massachusetts General from July 1977 to July 1979.	Morbidity and morality data; patterns of clinical practice.
SEARCH				
Rhode Island Health Services Research, Inc.	To provide a neutral resource organization whose health information sources could service a diverse need of all the State's health providers and planners, census and vital statistics, planning and utilization data.	Computerized data bases. State-wide cooperative reporting system to private, not-for-profit organization.	State of Rhode Island, since 1968.	Vital statistics; utilization rates, facility and expenditure data.
Drug Epidemiology Unit				
Boston University Medical Center	To study a broad range of problems concerning the clinical effects of drugs in humans, with particular emphasis on adverse effects, life-time histories of drug use; patients characteristics and diagnostic information.	Case-control surveillance system in 11 medical centers; nurse-monitors interview subjects.	Established July, 1976, over 10,000 individuals studied in United States and Canada.	Drug incidence and utilization data.
Pediatric Drug Surveillance Program				
Boston Children's Hospital Medical Center in collaboration with Drug Epidemiology Unit	To provide an estimate of the incidence of adverse reactions in hospitalized children.	Case-control surveillance; nurse monitors collection of information.	Established 1974; to date over 4,500 pediatric patients from premature newborns to young adults.	Drug incidence and utilization data.
Boston Collaborative Drug Surveillance Program (BCDSP)				
BCDSP/FDA	To provide some quantification of clinical efficacy and toxicity for prescribed drugs in specific types of patients.	Surveillance among hospitalized medical and surgical patients by specially trained monitors.	Since July 1966 in over 40 hospitals and seven countries on nearly 100,000 patients.	Drug incidence and utilization rates. Efficacy ratings of drugs used.
Dunedin Program				
Dunedin, Flordia Clinic	To screen participants over the age of 65 for medical disorders.	Data from patient records and interviews were collected.	Over 5,000 participants, an ambulatory geriatric population.	Drug incidence and utilization rates.
Olmstead Country, Minnesota System				
Mayo Clinic and Olmstead Medical and Surgical Group	To provide a complete medical record and information system for a circumscribed population for clinical and followup care; all contacts with the medical system are recorded.	Computerized medical records.	About 98 percent of Olmstead Country residents; medical records for population dating back to 1907.	Incidence and utilization patterns.
Seattle Group Health Cooperative System				
Seattle Group Health Cooperative	To develop comprehensive system tabulation on inpatient procedures and diagnoses, outpatient drug utilization enrollment features, and outcome measures.	Centralized automated data bank.	Over 270,000 members; 2 demographic base that resembles metropolitan Seattle, Wash.	Incidence and utilization rates; mortality, morbidity data; patterns of clinical practice.

System/organization	Purpose and focus of data	Method of data collection	Population/location/time coverage	Statistics generated
Health Services Research Center (Oregon Region) System				
Kaiser-Permanente Medical Care Program	To develop a computerized medical record for all center patients; extensive inpatient, outpatient, pharmacy and laboratory information.	Outpatient utilization records for a 5 percent sample of the Health Plan subscriber units and inpatient utilization records for 100 percent of hospital users.	Inpatient data over last 13 years.	Incidence and utilization rates; patterns of system and resource response
Manitoba Health Services Commission Databank				
Province of Manitoba, Canada	To provide for payment and control on population coverage; admission and discharge, service, diagnosis, surgical, followup information.	Comprehensive computerized databank through claims forms and individual hospital files.	All individuals registered in Manitoba (regardless of where care is received)	Registration, hospitals, and medical data
Seattle Heart Watch				
Communitywide registry	Angiographic data for determining the antecedents of sudden cardiac death.	Registration of all patients having coronary angiography and left ventriculography for systems or signs of ischemic myocardial disease.	A total of 2,616 patients from three private and two university teaching hospitals in Seattle, Washington from 1969 through 1974	Mortality and cardiovascular morbidity measures; 6-month interval followups

Appendix B.—Compendium of Bibliographic Data Bases for Medical Technology Assessment*

Data base/ subject	Producer	Online availability (United States)	Content	Sources of references	Coverage (earliest dates available)	Updating (approximate)
AGRICOLA (formerly CAIN) Agriculture, animal and plant sciences	USDA	NLM BRS, DIALOG, SDC	Citations to literature of agriculture.	Journals, monographs, government reports.	International—1970 to date.	12,000 per month.
AVLINE(AudioVisuals onLine) Science and technology (life sciences)	NLM		Descriptions of over 8,000 audiovisual and nonprint health science teaching materials.		Primarily United States—within the last 10 years.	200 records per month.
BIOETHICSLINE Science and technology (life sciences) bibliographic	Georgetown Kennedy Institute/Center for Bioethics	NLM	Citations to the literature on numerous moral, ethical, and policy issues of concern to the medical community.	60 indexes, 70 journals, other data bases. Corresponds to "Bibliography of Bioethics."	International—1973 to date.	450 per quarter.
BIOSIS PREVIEWS Science and technology (life sciences) bibliographic	BIOSIS	BRS, IRS, DIALOG, SDC,	Citations worldwide literature of research in life sciences. Original research reports, reviews of research, documentation, and retrieval information.	8,900 periodicals, books, monographs, conference proceedings, research communications, symposia are screened. Corresponds to "Biological Abstracts (BA)" and "Biological Abstracts/RRM."	International—1969 to date.	14,000 records from BA every month; 11,000 records from "BA/RRM" per month.
CA SEARCH; CASIA Science and technology (chemistry) bibliographic	Chemical Abstracts Service (CAS)	BRS, DIALOG, SDC	Citations to literature in chemistry: Journal articles, monographs, conference proceedings, technical reports, and patents.	Merger of Chemical Abstracts Condensates (bibliographic information from the printed "Chemical Abstracts)" and "CASIA (Chemical Abstracts Subjects Index Alert)".	International—1967 to date.	40,000 records per month.
CANCERLIT (formerly CANCERLINE) Science and technology (life sciences) bibliographic	National Cancer Institute, International Cancer Research Data Bank Program	NLM	Citations abstracts of literature on oncological epidemiology, pathology, treatment, and research.	Primary literature: articles selected from over 3,500 journals, monographs, technical reports, conference proceedings, and theses.	International—1963 to date.	About 4,000 citations per month.
CANCERPROJ (Cancer Research Projects) Science and technology (life sciences)	Current Cancer Research Project Analysis Center operated by Smithsonian Science Information Exchange (SSIE)	NLM	Summaries of ongoing and recently completed cancer research: U.S. Federal and non-Federal SSIE.		International—past 2 to 3 fiscal years.	Quarterly, average size is 18,000 projects.
CHEMDEX; CHEMLINE; CHEMNAME™; CHEMX Science and technology (chemistry) properties	Chemical Abstracts Service (CAS)	DIALOG, SDC, NLM	Chemical dictionaries based on CAS Registry Nomenclature File, nomenclature, synonyms, structural data, molecular formula, ring system information.			Varies, from quarterly to less frequently.

---

*As of December 1981.

Data base/ subject	Producer	Online availability (United States)	Content	Sources of references	Coverage (earliest dates available)	Updating (approximate)
CHILD ABUSE AND NEGLECT						
Social sciences (bibliographic and referral)	Department of Health and Human Services, National Center of Child Abuse and Neglect	DIALOG	Citations and abstracts of materials definition, identification, prevention, and treatment of child abuse and neglect. References to the literature audiovisual materials, excerpts of current laws listings of U.S. service programs.	"Child Abuse and Neglect Research: Projects and Publications, Child Abuse and Neglect Programs" and "Child Abuse and Neglect Audiovisual Materials."	United States—1965 to date.	Research projects, audiovisual materials and legal citations, annually; literature, semiannually.
CLINPROT (Clinical Cancer Protocols)						
Science and technology (life sciences)	National Cancer Institute, International Cancer Research Data Bank Program (ICRDB)	NLM	Summaries of clinical investigations on new anticancer agents and treatment modalities.		International—1978 to date.	Quarterly; several hundred records per year.
CDI (Comprehensive Dissertation Index)						
Multidisciplinary	University Microfilms, Inc.	BRS, DIALOG, SDC	Citations to all dissertations accepted for Ph.D. at accredited U.S. and 210 foreign institutions.	Corresponds to printed "Dissertation Abstracts International (DAI)" and "American Doctoral Dissertations (ADD)".	International—1961 to date.	3,000 "DAI" per month and 3,000 "ADD" per year.
CPI (Conference Papers Index)						
Science and technology general	Cambridge Scientific Abstracts	DIALOG, SDC	Citations to papers presented at 800 scientific meetings.	Corresponds to printed "Conference Papers Index".	International—1973 to date.	8,300 records per month.
DRUGINFO and ALCOHOL USE/ABUSE						
Social sciences (bibliographic)	University of Minnesota, College of Pharmacy, Drug Information Service Center	BRS	Two files on alcohol and drug use/abuse citations to monographs, journals, conference papers, instructional guides, films on aspects of alcohol and drug use/abuse. Research in the area of chemical dependency.		U.S. DRUGINFO, 1968 to date; ALCOHOL USE/ ABUSE, 1968 to 1979.	Quarterly.
ENVIROLINE						
Science and technology (environment)	Environment Information Center	DIALOG, SDC	Citations (and abstracts from 1975) topics related to environment.	Corresponds to printed "Environmental Abstracts".	International—1971 to date.	700 per month.
EPILEPSYLINE						
Science and technology (life sciences) bibliographic	National Institutes of Health, National Institute of Neurological and Communicative Disorders and Stroke	NLM	Citations and abstracts of the literature on epilepsy. Covers basic sciences; seizures etiology; genetics; systemic changes related to seizures; diagnostic aids; psychology, sociology, and epidemiology.	Excerpta Medica publication, "Epilepsy Abstracts".	International—1945 to date.	400 records per quarter.

Data base/ subject	Producer	Online availability (United States)	Content	Sources of references	Coverage (earliest dates available)	Updating (approximate)
EXCERPTA MEDICA; EMBASE Science and technology (life sciences) bibliographic	Excerpta Medica	DIALOG	Citations and abstracts of worldwide biomedical literature on medicine and areas of biological sciences related to medicine, including clinical practice, research, and economic and management issues.	References to articles from over 3,500 domestic and international journals. Corresponds to the 43 abstract journals and 2 printed indexes that comprise the printed "Excerpta Medica." Approximately 100,000 citations are added to the data base each year that do not appear in the printed publication.	International—1969 to date.	5,000 records weekly.
FOODS ADLIBRA Science and technology (food science) bibliographic	Komp Information Services, Inc.	DIALOG	Citations and abstracts to the journal literature on food technology nutritional and toxicological information.		1974 to date.	2,000 records per month.
HEALTH INDEX INFO Health statistics (bibliographic and referral)	NCHS, Clearinghouse on Health Indexes	NCHS	Citations to unpublished literature relating to measures of health status and quality of life: Journal articles, monographs, conference proceedings, and technical reports.	Reports of ongoing research, technical reports, workshops, conference proceedings, and symposia as well as the MEDLINE data base, current contents, and selected periodicals.	International—1973 to date.	300 per quarter.
HEALTH (Health Planning and Administration) Science and technology (life sciences) bibliographic	NLM	NLM, BRS	Topics relevant to the provision of health care services.	3,000 journals in MEDLINE, journals reviewed for "Hospital Literature Index," other monographs, technical reports. Forthcoming will include citations from CATLINE, "Health Planning Series" published by NTIS.	International—1975 to date.	2,000 records per month.
HISTLINE Science and technology (life sciences) bibliographic	NLM	NLM	Citations to articles, books, conference proceedings and other literature on the history of medicine.	The MEDLINE data base and "Current Catalog." Corresponds to NLM's annual "Bibliography of the History of Medicine."	International; some data from 1964.	Quarterly.
IPA (International Pharmaceutical Abstracts) Science and technology (pharmaceuticals) bibliographic	American Society of Hospital Pharmacists	DIALOG	Citations and abstracts of the literature on development and use of drugs and to clinical, practical, theoretical, scientific, economic, and ethical aspects of professional pharmaceutical practice.	Corresponds to the printed "International Pharmaceutical Abstracts."	International—1970 to date.	About 1,200 records every 2 months.
IRL LIFE SCIENCES COLLECTION Science and technology (life sciences) bibliographic	Information Retrieval, Ltd. (IRL)	DIALOG	Citations and abstracts to worldwide life sciences literature.	Corresponds to the 15 abstracting journals published by IRL.	International—1978 to date.	Monthly.
LADB (Laboratory Animal Data Bank) Science and technology (life sciences) numeric	Battelle Columbus Laboratories, under contract to NLM	NLM	Numeric data base system contains comparative data on control animals used in biomedical research.	Data are collected from Government agencies, independent laboratories, pharmaceutical manufacturers, universities, and animal producers.	United States.	Monthly.

Data base/ subject	Producer	Online availability (United States)	Content	Sources of references	Coverage (earliest dates available)	Updating (approximate)
MEDLINE (MEDLARS onLine)						
Science and technology (life sciences) bibliographic	NLM	BRS, NLM, DIALOG	Citations and sometimes abstracts of worldwide biomedical literature on research, clinical practice, administration, policy issues, and health care services.	References to articles from 3,000 journals published in the United States and 70 other countries, chapters and articles from selected monographs. Author abstracts (from 1975) are available for some citations. Corresponds to coverage of printed "Index Medicus."	International, NLM, 1966 to date (current year and two preceeding years are searchable online, earlier years are searched offline).	About 20,000 records per month.
MEDOC						
Science and technology (life sciences) bibliographic	University of Utah, Spencer S. Eccles Library	BRS	Citations to health-related U.S. Government documents: medical and health sciences, Food and Drug Administration, mental health, biomedical engineering. Safety; child development; juvenile delinquency; welfare; Medicare and Medicaid; aging; rehabilitation; alcoholism; ethnic groups; basic sciences; nutrition; veterinary medicine; behavioral sciences.		United States—1976 to date.	Quarterly.
NIMH (National Institute of Mental Health)						
Science and technology (life sciences) bibliographic	National Institute of Mental Health, National Clearinghouse for Mental Health Information (NCMHI)	BRS, DIALOG	Citations and abstracts of the mental health literature, both the biomedical and social aspects.	Materials are from 1,000 journals, books, technical reports, workshops, conference proceedings, and symposia.	International—1969 to date.	About 3,500 records per month.
NTIS (National Technical Information Service)						
Science and technology (general) including biology (bibliographic)	National Technical Information Service	BRS, DIALOG, SDC	Citations, abstracts of technical reports U.S. and non-U.S. Government-sponsored research. Announcements of computer-readable software data files, federally sponsored translations.	"Government Reports Announcements and Index (GRA&I);" in part to the weekly "Abstract Newsletters." Over 750,000 citations.	United States—1964.	2,600 records every 2 weeks.
POLLUTION						
Science technology (environment)	Cambridge Scientific Abstracts	BRS, DIALOG	Citations and abstracts to worldwide literature on pollution research.	Corresponds to printed "Pollution."	International—1970 to date.	1,400 records per 2 months.
POPLINE						
Social sciences demographics	Johns Hopkins University	NLM	Citations, most with abstracts to literature, on demography, contraception, fertility.	Worldwide literature.	International—1970 to date.	Monthly.
POPULATION BIBLIOGRAPHY						
Social sciences (demographics and population) bibliographic	University of North Carolina, Carolina Population Center	DIALOG	Literature on population research abortion, demography, family planning, fertility policy and research methodology. Monographs, journals, technical reports, government documents, conferences, unpublished reports. Socioeconomic as opposed to biomedical aspects.	Worldwide literature.	Primarily United States—1966 to date.	800 records every 2 months.

Data base/ subject	Producer	Online availability (United States)	Content	Sources of references	Coverage (earliest dates available)	Updating (approximate)
PRE-MED						
Science and technology (life sciences) bibliographic	BRS	BRS	Articles in 125 key medical journals "Abridged Index Medicus" appear within 10 days of the appearance of the journals in major health sciences, libraries, clinical medicine, nursing, and hospital administration.	Biomedical literature from the United States, Canada, Great Britain.	United States, Canada, Great Britain, most current 3 to 4 months.	Weekly.
PSYCHOLOGICAL ABSTRACTS; PSYCHABS; PSYCINFO						
Social sciences(psychology) bibliographic	American Psychological Association	BRS, DIALOG, SDC	Citations and abstracts of journals, psychology, the behavioral sciences (human and animal).	1967 to 1979, corresponds to the printed "Psychological Abstracts." In 1980 the data base contains more references than the printed publication.	International—1967 to date.	About 2,600 records per month.
RINGDOC						
Science and technology (pharmaceuticals) bibliographic	Derwent Publications, Ltd.	SDC	Citations to the worldwide journals on pharmaceuticals, chemical literature related to clinical medicine.	"RINGDOC Abstracts Journal."	International—1964 to date.	4,500 records per month.
RTECS (Registry of Toxic Effects of Chemical Substances)						
Science and technology (chemistry) properties	National Institute for Occupational Safety and Health (NIOSH)	NLM	75,000 toxicological measurements of 40,000 chemicals.	Corresponds to the printed and microfiche "RTECS" publication available from the U.S. Government Printing Office.		Quarterly.
SAFETY						
Science and technology (safety) bibliographic	Cambridge Scientific Abstracts	SDC	Citations and abstracts of journals, patents, dissertations, government reports, corporate research reports, conference proceedings: general, occupational, transportation, aviation, aerospace, environmental, ecological, medical safety.	Corresponds to the printed "Safety Science Abstracts Journal."	**International—June 1975 to date.**	3,000 records every 2 months.
SCISEARCH						
Science and technology (general)	Institute for Scientific Information	DIALOG, BRS	**Citations to the worldwide literature wide range of scientific technological disciplines.**	300 journals selected items from "Current Contents" publications. Corresponds to printed "Science Citation" Index.	International—1974 to date.	52,000 records per month.
SERLINE (SERials onLine)						
Science and technology (life sciences) bibliographic	NLM	NLM	Citations to approximately 30,000 serial titles, on order, in process, or currently received at NLM. One fifth of the records contain locator information.	**Corresponds to the microfiche publication, "Health Sciences Serials."**	International—1965 to date.	Monthly.
SOCIAL SCISEARCH						
Social science and humanities	Institute for Scientific Information	BRS, DIALOG, SDC	Citations to 1,500 social science journals and 2,900 journals in related fields.	Corresponds to printed "Social Science Citation Index."	International—1972 to date.	11,700 records per month.

Data base/ subject	Producer	Online availability (United States)	Content	Sources of references	Coverage (earliest dates available)	Updating (approximate)
SSIE; SSIE Current Research Multidisciplinary (general) including biomedical science		BRS, DIALOG, SDC	Descriptions, references to research in progress and recently completed research sponsored primarily by federal government. Basic and applied research; life, physical, social and behavioral, engineering sciences.		Primarily United States—1974 to date.	About 9,000 records per month.
TDB (Toxicology Data Bank) Science and technology (chemistry) properties	NLM, Toxicology Information Program	NLM	Data on 1,700 substances of known or potential toxicity; nomenclature information, properties, pharmacological, toxicological, environmental and manufacturing information.	Extracted from the published literature reviewed by subject specialists.		200 new records (i.e., additional substances).
TOXLINE Science and technology (chemistry) bibliographic	NLM, Toxicology Information Program	NLM	Citations and abstracts to the worldwide literature in toxicology.	Comprises 11 subfiles on various aspects of toxic substances.	International. Some information from 1940; varies according to file.	About 8,000 records per month.
VETDOC (Veterinary Literature Documentation) Science and technology (animal science) bibliographic	Derwent Publications Ltd.	SDC	Citations to the worldwide literature veterinary drugs, vaccines, toxicology.	Corresponds to "VETDOC Abstracts Journals."	International—1968 to date.	About 1,000 records per quarter.

Appendix C.—Assessment of Medical Technology: Methodological Considerations

by Paul M. Wortman, Ph. D., University of Michigan
and
Leonard Saxe, Ph. D., Boston University*

Abstract

This appendix is primarily concerned with methodological issues underlying the research evidence used to assess medical innovations. In particular, it examines the process of research analysis in interpreting the results from individual studies and the complementary process of research synthesis in aggregating the results from many studies. Both processes are important to medical technology assessment and require an understanding of their methodological limitations. A conceptual framework is presented for determining the validity of the research evidence derived from various methodologies (e.g., clinical trials, consensus exercises) employed to assess medical technology.

Introduction

Medical technology has assumed an increasingly central role in the delivery and costs of health services. In order to assess the effectiveness of medical technologies and increase the impact of Federal funds, Congress has undertaken a number of policy initiatives over the past few years (310). Through the 1976 Medical Device Amendments, it expanded the Food and Drug Administration's (FDA's) role in assessing medical products for safety and effectiveness. In 1978, it established the National Center for Health Care Technology (NCHCT) with a mandate to conduct medical technology assessments. Technology assessment has been defined as a "comprehensive form of policy research that examines the . . . social consequences of technology" (7,269). Technology assessments must consider a wide range of outcomes of a technology, including safety, efficacy, cost effectiveness, and social impact. These outcomes are judged by considering various forms of information about a technology. This information is typically derived from multiple studies that vary in their methodological adequacy and appropriateness to assess the technology.

Despite their importance, methodological features of the research studies used to develop a technology assessment are often given only minimal attention. Methods that have very different functions and applicability, such as controlled clinical trials and consensus development, are often lumped together and viewed as alternatives to one another (see 266). Similarly, randomized clinical trials (RCTs) are often seen as a unitary method, although they represent a diverse set of procedures. In addition, while the usefulness of a technology assessment rests on the ability to integrate research evidence, little attention is paid to research synthesis activities. There are no clear-cut standards for the quality of evidence that should be considered nor for the ways in which discrepant information should be consolidated. Recent Government conferences on methods for assessing medical technology have not changed the situation (e.g., 3,87).

The pressures for diffusion of medical innovations require valid statements of efficacy, safety, and social impact (266). These, in turn, necessitate appropriate methods for assessment. Proper research methods allow one to state with confidence that observed effects are actually due to the medical innovation—i.e., well-designed and carefully conducted evaluative studies for technology assessments will produce valid and reliable results. As the remainder of this appendix will demonstrate, the failure to conduct proper studies often results in serious criticism of both the validity of the research and the validity of the technology assessments based on this research. A framework for determining validity that can be used to interpret the results of individual studies and to synthesize the findings from many studies will be presented.

The purpose of this appendix is to review some principles for interpreting and integrating the results of evaluative studies that underlie the assessment of medical technologies and to indicate their place in a general strategy for medical technology assessments. The remainder of this appendix is organized in three sections. The section immediately below discusses the interpretation of individual evaluative studies of medical technology. It introduces validity concepts and describes the relationship between the design of research studies and the usefulness of the information generated. Three broad categories of designs are discussed: 1) RCTs; 2) controlled clinical trials lacking randomization, also known as quasi-experiments (45);

**Authors' acknowledgments:* The authors would like to thank Drs. Fred Bryant, John McSweeny, William Yeaton, and the many anonymous reviewers for their helpful comments on previous drafts. The authors also thank Marga Van Goethem and Jean Holther for the many hours of conscientious typing.

The authors accept full responsibility for any remaining infelicities of prose and inaccuracies of thought. The views expressed in this paper are those of the authors and do not necessarily represent the views of OTA.

and 3) uncontrolled studies known as nonexperimental investigations or case studies. Some typical designs are described, and the problems they pose in interpreting the evidence from studies assessing health care technology are presented. The second section below examines methods for synthesizing the results from many studies. These include formal quantitative procedures (e.g., meta-analysis) and group decisionmaking techniques (e.g., consensus conferences). The validity problems in using these procedures are discussed. The final section briefly describes a strategy for integrating these assessment methods with the innovation process.

Research Analysis: Interpreting the Results of Individual Studies

A thorough technology assessment is viewed as including 10 elements (269). One of the most important elements in a technology assessment is the "evaluation of potential impacts," which encompasses "technical feasibility" (i.e., effectiveness), safety, ethics, and economic considerations. If technology assessments are to be useful, their evaluation component must be conducted in a systematic manner that employs acceptable scientific methods, especially research design (60,144). Proper research design is of utmost importance if the observed changes in a patient population are to be correctly attributed to the technology being assessed rather than to some extraneous factors. As the following discussion will show, it is often these other unrelated factors that cloud the interpretation of technological impact and undermine the validity of the technology assessment.

Validity

Validity involves the careful analysis of research to determine its adequacy or scientific soundness. The analysis of research requires an understanding of the strengths and weaknesses of the methods used to generate scientific evidence. The problems in determining the validity of the findings in research studies have been of continual interest in medicine. Recently, the medical journal *Lancet* (170,171) carried a series dealing with research design issues in "assessing clinical trials." The articles discussed problems that can undermine the validity of clinical trials, especially those dealing with medical innovations. A useful conceptual framework for examining issues of validity has been developed by Cook and Campbell (68). These methodologists organize validity problems into four categories: 1) internal validity, 2) statistical conclusion validity, 3) external validity, and 4) construct validity. These four categories provide a useful way of understanding the implication of design issues for medical technology assessment studies.

INTERNAL VALIDITY

Internal validity refers to whether the observed effects of a medical innovation are truly due to the technology and not to some other factors. Internal validity, therefore, is the most important component of validity. An important part of any technology assessment asks questions such as: Would patients have improved even if they did not receive the innovation? or, Do they really improve more with the innovative procedure than with the traditional approach? An evaluative study that can adequately answer these questions is called an internally valid evaluation. From a scientific perspective, internal validity involves the assignment of causality to the innovation for the observed benefits or risks.

A key issue in the internal validity of an assessment is the "control" of factors extraneous to the innovation. When random assignment of patients to treatment and control groups fails or is not employed, a number of plausible alternative explanations can be offered. These so-called "threats to validity" include, among others, alternative hypotheses based on selection and statistical regression (5,68). *Selection* or selection bias occurs when patients are assigned to receive a treatment because of particular characteristics (e.g., better prognosis), while *statistical regression* arises when patients are chosen because of their extreme value on a laboratory test or other measure relevant to the treatment. Many of these validity threats are defined, described, and discussed in the following sections. Their usefulness in interpreting the evidence from technology assessment studies will be demonstrated.

A properly conducted RCT is internally valid. Even when studies are advertised as RCTs, however, one should carefully examine their methods or procedures to determine if the randomization process was properly conducted. A recent RCT published in the *Journal of the American Medical Association* by Hoehler, et al. (190), illustrates the problem. To assess the effectiveness of a rotational spinal manipulation for back pain, the authors report, 95 subjects were admitted to the trial and were "randomly assigned to either the experimental or the control group." From this brief description of the randomization procedure, one would expect that 45 to 50 subjects would be assigned to each condition. Instead, the initial table reveals that there were 56 in the experimental, spinal manipulation condition and 39 in the control group. This appears to be quite divergent from what a randomized process would produce. (The probability of this difference occurring is greater than the "1 in 20" level associated with chance.) Although there may be good reasons for this discrepancy besides chance, the authors are mute on this point. One is left with the suspicion that other factors (e.g., severity of pain) may have influenced patient assignment to conditions ar. that these selection factors may be responsible for the observed results. These factors would pose a threat to internal validity due to differential patient selection into the two groups.

STATISTICAL CONCLUSION VALIDITY

There are many threats to the validity of a technology assessment study. Threats related to the analysis of the data are particularly important, and Cook and Campbell have called these threats to statistical conclusion validity. This category of validity focuses on the appropriateness of statistical tests and their ability (or power) to determine whether or not observed effects are due to chance. Many, otherwise internally valid, studies in health have used too few subjects (see 153,171) to detect anything but the largest effects. Statisticians call this a Type II error—the acceptance of a finding of no difference (in effectiveness) when it is false. It is possible that some useful technological innovations have been discarded due to faulty statistical procedures. For example, a recent study (264) on the effectiveness of timolol in reducing mortality after a heart attack noted that one reason most other studies of these beta-blockers have found little or no effect was that they contained too few patients "to exclude the possibility that a beneficial effect was being overlooked."

EXTERNAL VALIDITY

External validity concerns the generalizability of the observed effects to other patient populations, settings, or conditions. That is, would the treatment be beneficial in other settings or are its effects specific to the present situation? The concept of external validity is captured in OTA's definition of "efficacy" (266), the likelihood of benefit under optimal circumstances to "individuals in a defined population . . . ." The importance of external validity considerations can be found in an example drawn from the first National Institutes of Health (NIH) (96) consensus development conference on the efficacy of mammography in the detection of breast cancer. The panel concluded that the technology was only beneficial for women over 50 and might be harmful for others due to the risks of repeated exposure to radiation. These conclusions, based largely on one study (341), indicated dramatic differences in effectiveness from one subpopulation of women to another.

CONSTRUCT VALIDITY

The last type of validity deals with conceptual issues. It depends on the adequacy of the theory that one has about what makes the innovation effective and the adequacy of the measures of the observed effects (or variables) derived from the theory. The recently concluded debate on the efficacy of radical mastectomy demonstrated the role of theory (147). Once it was shown that cancer was disseminated through the blood stream, the basis of the Halsted radical surgery was called into question. Construct validity also refers to improper measurement of outcomes as well as improper control of the technology. The latter can often be confused or contaminated by other changes that may cause the observed effects.

Outcome Measures.—One of the major problems in assessing medical technology is the absence of good outcome measures of the constructs considered important. For example, a researcher in behavioral medicine attending a conference on the social impact of coronary artery bypass graft (CABG) surgery noted, "There is no consensus about how you define and measure quality of life" (284). As a consequence, technology assessment studies often focus on a variety of process variables (e.g., admissions, length of stay, etc.) that may, or may not, be indicative of the delivery of services and are not concerned with the overall impact of a technology on patient health. Often, the absence of such observations can be traced to the lack of a specific, well-defined treatment procedure.

Even where there are outcome measures or end points, these may be "soft"[1] or subjective. Relief of angina in CABG surgery is a case in point. Both patients' and physicians' expectations concerning the benefits of surgery (see 297) may influence judgments of relief. Such expectations are the rule for the technological advances in modern medicine. As discussed, it is essential to eliminate from technology assessment studies the potential bias produced by these expectations of efficacy (i.e., placebo effects). Good measures of the impact of innovative treatments are needed. Often, the debate on the efficacy of medical technologies swirls about very few objective outcome measures (e.g., survival in CABG surgery, cesarean section in fetal monitoring).

---

[1] Paul Meier, University of Chicago, personal communication, December 1980.

Design Categories

Several types of designs have been used in studies evaluating or assessing medical technologies. These generally fall into the three categories noted above: 1) RCTs, 2) quasi-experiments or controlled trials, and 3) uncontrolled case studies. These designs represent the principal methodological approaches to research studies of medical technology assessment and vary with respect to the validity of the evidence they produce. In this section, the advantages and disadvantages of each design category are examined with respect to validity. The major concern in this discussion is with the choice of an appropriate control or comparison group and the effect this has on the validity of the findings.

RANDOMIZED CLINICAL TRIALS

The "true" experiment or RCT is the preferred design for producing unambiguous assessments of a medical technology (see 60,187). The essential ingredient in an RCT is randomization: Patients or other experimental units are randomly assigned to experimental (treatment) or control conditions. Although some (76) argue that only posttreatment measurement of patients is required in an RCT, most health researchers use both pretreatment and posttreatment measures. This provides a check on the initial or baseline equivalence of the groups and an accurate (or unbiased) estimate of the amount of change produced by the innovation. The basic question asked in a true experiment is whether effects observed in the experimental (or treated) group are also observed in the control (or untreated) group. If the answer is essentially "no," the effects may be safely attributed to the technology.

RCTs are in reality a family of designs that vary in size and complexity. The number of treatment conditions can vary (e.g., dosage levels) as can the size of the population and the theoretical significance of the study. Small randomized trials are often performed early in the development of a technology to demonstrate or test the efficacy of the treatment's innovative elements. Such studies typically involve only a single investigator observing a few subjects—either animals or humans—at a single site. At another level, large-scale, multicenter trials are often conducted to establish the efficacy or safety of a developed technology. Such RCTs are usually necessary to provide the appropriate number and type of patients to assess the technology as quickly as possible. Moreover, the diversity of sites and subjects can provide useful data on the external validity of the innovation. These multicenter trials are not immune from problems. They add difficulties in organizational complexity and hence limit the researcher's ability to assess the technology under "ideal conditions." Indeed, much of the debate over the Veterans Administration's (VA's) multicenter RCT of CABG surgery centered on such problems (254,255). There were wide differences both in types of patients selected and in the operative mortality among the sites. Conner (66) discusses several strategies for successfully maintaining the integrity of the randomization process (e.g., a centralized procedure with few implementers).

Blinding.—Other attributes of RCTs are also important to note. In order to reduce the bias in physician and patient expectations, physicians and patients should both be unaware of or "blind" to the treatment the patient is receiving. This is called a "double-blind study" and is frequently used in assessing drugs. In some cases, such control is not possible. For example, today it would be ethically impossible to give some patients sham surgery to assess the efficacy of CABG surgery or even to give them the much simpler internal mammary artery ligation surgery that was proven ineffective using such a control group (20). And even if it were possible, only the patients would not know which "treatment" they had received (i.e., the study would be a single-blind study). As noted above, the inability to blind patients and physicians contributes to construct validity problems in interpreting the surgery's effect on the relief of angina.

When researchers and patients are not blind to the treatment being delivered, it is possible that their expectations can affect (or be confounded with) the outcomes. To avoid this, RCTs often use a placebo (i.e., a procedure that appears identical to the innovation but has no therapeutic benefit). A good example of a relatively uncomplicated RCT employing a placebo is provided by The Coronary Drug Project (72) that assessed the effectiveness of a drug using this technique. Placebo control is useful in establishing construct validity but is hard to employ with most non-drug innovations.

The Hoehler, et al. (190), study of spinal manipulation for relief of back pain (noted above) is an exception in that it employed a placebo treatment for an assessment of a "technique." The control group patients received a "soft-tissue massage of the lumbrosacral areas." The authors assumed that this was a valid placebo because their previous research showed that "patients with no knowledge of spinal manipulation probably cannot distinguish that therapy from soft-tissue massage." They found no significant difference at discharge as both groups were substantially improved. In fact, the observed "dramatic" effects of a number of innovations (e.g., gastric freezing and internal mammary artery ligation) were later shown by well-designed RCTs to be due to a "placebo effect."

Chalmers, et al. (54), maintain that research investigators must also be blind to the randomization process and to the interim results while the trial is in progress. In the former situation, bias could affect patient assignment; in the latter, it could also affect patient withdrawals. In either case, the validity of the study is jeopardized.

One should be cautious, however, in assuming that all RCTs are necessarily exemplary and immune to threats to their validity. Research reports often conceal major flaws in the conduct of the RCT. The recent critique of the Anturane Reinfarction Trial by FDA (363) provides a cogent illustration of the problems that can occur. The FDA audit found major errors in coding outcomes and classifying patients that undermine the credibility of the study. For example, errors made in the assignment of cause of death systematically favored finding a benefit for sulfinpyrazone (Anturane) in reducing mortality following myocardial infarction. Moreover, the classification scheme itself was found to be lacking in meaning (i.e., in construct validity).

Statistical conclusion validity is also important in assessing RCTs. Most of the RCTs on coronary bypass surgery have data analysis problems stemming from serious attrition (or *experimental mortality)* in the medically treated condition, with patients crossing over into the surgery group. The various analytic approaches for handling this problem have been inadequate (400), including those based on initial patient assignment or "intention-to-treat" (289). A reexamination of the crossover problem indicates that such inappropriate statistical analyses may result in a Type II error. If the worst medical cases are switching (as is indicated), then the mean outcome for their group is being inflated. For example, a simple algebraic calculation indicates that the observed amount of crossover (i.e., one-sixth) by the worst medical patients would increase the mean by at least one-fifth of a standard deviation (i.e., 0.2 SD) or 20 percent. Since survival data usually have a negatively skewed distribution, the increase in the mean could be more. Thus, it is possible that the crossovers in the coronary bypass RCTs conceal a surgically significant difference larger than 25 percent between the two groups.

Conclusions.—Although it is often true that large-scale randomized experiments are more expensive to conduct and require more planning than nonexperimental designs (see 222), that is not always the case and they should not be rejected out of hand. Reviews of health research practices indicate that the use of inexpensive nonrandomized designs often produces costly errors, since faulty results can lead to incorrect conclusions and inappropriate policy decisions (42,158, 335). Gastric freezing provides a classic example (143). Hundreds of devices were purchased by physicians based on the evidence from poorly designed, nonrandomized studies using few patients. In many cases, the greater confidence in the results of an assessment that an RCT permits greatly outweighs any difficulties in its implementation.

CONTROLLED CLINICAL TRIALS

Despite the advantages of randomized experiments, they are often difficult to implement in settings such as hospital clinics and physicians' offices. McKinlay[2] has pointed out that RCTs are especially difficult to conduct for existing technologies that are already widely diffused. Unfortunately, widespread diffusion has been a frequent occurrence in assessing medical innovations (see 253). In such situations, administrators are usually reluctant to make the changes in policies and procedures needed to conduct a randomized experiment. Another important obstacle to conducting RCTs that is common in health evaluations is the a priori conviction of medical personnel that specific patients are best suited for the innovative treatment being evaluated. In this case, staff will resist and possibly even subvert the randomization process. For example, the assessment of high-oxygen environments as a cause of retrolental fibroplasia in premature infants was impeded by well-intentioned nurses (346). In one study, nurses raised the oxygen level for the experimental group babies in the belief that the low-oxygen environments were harmful. In another study, it was necessary to implement the treatment only partially, until evidence of the harmful effects of oxygen were more apparent. The corruptive behaviors derived from preconceived attitudes pose an additional barrier to conducting an assessment study in an applied field setting.

Sometimes researchers find that conditions prohibit RCTs. This problem can occur for a variety of reasons: politics, as noted below in the Salk Vaccine Trial; corruption of the design through attrition or other implementation problems; ethical prohibitions where patients or physicians have been persuaded of the efficacy of a treatment (see 171); or cost considerations where funds for a long-term local study are unavailable. Sometimes, unfortunately, nonrandomized studies are conducted because of a naive belief in the ability of statistical techniques to correct for the biases introduced by selection.

When randomized experiments are not feasible, investigators often use one of several quasi-experimental designs (see 45). Quasi-experiments involve the use of self-selection procedures in the assignment of patients to either treatment or control conditions. These designs do not permit the rigorous controls provided by RCTs. Even good quasi-experiments allow some competing explanations for observed treatment effects. In particular, two quasi-experimental designs—the cohort design and the time-series design—are commonly used. The validity of these designs is discussed below.

---

[2] J. B. McKinlay, "From 'Promising Report' to 'Standard Procedure': Seven Stages in the Career of a Medical Innovation," *Milbank Mem. Fund Q.* 59:374, 1981.

Cohort Design.—The cohort study or nonequivalent control group design (NECGD) is the quasi-experiment that results when random assignment of subjects to the treatment and control conditions is not employed (see table C-1). Because random assignment is not used, the "treatment" and "control" groups are "nonequivalent" and may differ in systematic ways. In the discussion that follows, the term "comparison group" is used instead of "control group" when that situation obtains.

Roos, et al. (319), employed a cohort or NECGD to determine the effectiveness of tonsillectomy with or without adenoidectomy. Using claims and patient registration data provided by the Manitoba Health Services Commission, the investigators were able to create two comparison groups to assess the impact of these surgeries on subsequent episodes of respiratory illness. The first, and larger, group consisted of operated and nonoperated persons under the age of 14 covered during a 3-year period, whose records indicated evidence of tonsillar illness. For the experimental (operated) group there had to be data available for 1 year before and 1 year after their surgery. The records of the comparison group had to indicate that they remained unoperated during this period.

A number of threats to the internal validity of this study were examined by Roos, et al. (319). Since both treatment and comparison groups were similar in age and sex, it was felt that *maturation* (i.e., changes in health with age) was not a threat to validity. Moreover, by using concurrent controls, *history* (i.e., the effect of temporal events such as new health practices) was also eliminated as a threat. However, *'local' history* (68) (i.e., dealing with familial or physician factors such as predisposition toward surgery) may have differed among the two groups. To reduce the influence of this potential threat, a second comparison group, composed of the siblings of those operated on, was also used.

Table C-1.—Nonequivalent Control Group or Cohort Design

	Pretest		Posttest
Treatment group	0	X	0
	—R		
Comparison group	0		0

0 = Observation or measurement
X = The application of the treatment or technology
—R = Absence of randomization

SOURCE: Adapted from T. D. Cook, and D. T. Campbell, *Quasi-Experimentation: Design and Analysis of Research in Field Setting*, 1979.

Statistical conclusion validity (i.e., the correctness of the data analysis) was also strengthened by using two different analytic approaches as well as subsidiary analyses to eliminate effects due to statistical regression. This latter threat was examined by stratifying the two groups according to the number of preoperative episodes of respiratory illness. If regression was causing or influencing the results, there would be greater changes in the persons with the most preoperative episodes (i.e., the extreme scores). In all cases, the results were the same—i.e., the operated group showed (statistically significant) fewer postoperative cases of respiratory illness. Additional analyses to control for the severity of the illness (by examining specific diagnostic categories) yielded similar results.

The findings of this study are not meant to be definitive with respect to the effectiveness of tonsillectomy. This procedure has become the focus of some debate with the advent of antibiotics (389), and RCTs are currently being conducted. However, the study is an instructive methodological example that illustrates the assessment of actual practice or the "effectiveness" of an innovation.

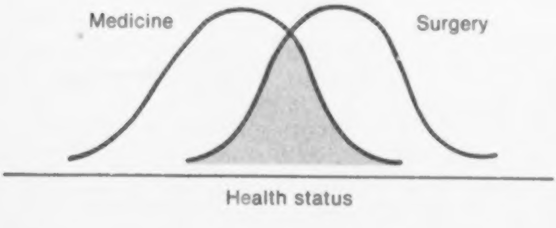
Matching.—One common form of the cohort design involves examining naturally occurring patient populations (as in the above example on tonsillectomy) to determine whether they differ on important characteristics and then statistically adjusting for these differences. These procedures are referred to as matching or retrospective matching. Two examples from the literature on CABG surgery illustrate the approach and its problems.

McNeer and his associates (240) examined the data drawn from 781 consecutive patients treated for coronary artery disease at Duke University Medical Center between 1969 and 1973. Of these patients, 402 were treated medically and 379 had bypass surgery. Patients were compared on 89 baseline variables. The authors believed that "therapeutic decisions tend to be random." They found the two groups to be "remarkably similar" and the results to be unchanged when individual variables were corrected or statistically adjusted for initial differences. However, as Ross (326) noted in his review of this study, there was a systematic pattern of differences among significant variables such that the "surgical cohort would have a better prognosis irrespective of the form of therapy . . . ." For example, surgical patients had (statistically significant) more positive exercise tests, higher ejection fraction, and smaller heart size. The separate analyses and adjustments do not correct for this systematic bias, and the conclusions are therefore rendered suspect. It is possible that the outcomes merely reflected the preexisting differences among the two groups.

One common form of this design involves the direct matching of patients receiving different therapies or treatment. A recent example of this method is the study conducted by Hammermeister, De Rouen, and Dodge (181) to assess the efficacy of CABG. Data from the Seattle Heart Watch angiography registry were used to form 287 matched pairs of surgical and medical patients. The patients were matched on seven variables (e.g., ejection fraction, arrhythmia, number of stenotic arteries). An analysis of the actuarial survival rates resulted in a statistically significant finding indicating decreased mortality for patients treated surgically. When the data were analyzed by the amount of coronary disease (i.e., one-, two-, or three-vessel disease), improved survival due to surgery was detected only in the subgroup of 97 pairs with two-vessel disease.

Campbell and his associates (43,44) have graphically demonstrated the problems posed by a matching design. In particular, they note that *statistical regression* to the mean (usually abbreviated as "regression") is a major threat to the internal validity of the results in such designs. The basic regression phenomenon can be easily illustrated. Assuming that the two groups (in this case, medical and surgical patients) differ on some relevant unmeasured variables, it is possible that they may be drawn from populations that differ in their health status (see fig. C-1). Given that surgeons are likely to select the best candidates for this procedure, the assumption seems warranted. The resulting matching procedure would then pair medical patients above their group's mean with surgical patients below their group's mean. Given the imperfect (or unreliable) measures used, the two groups will regress to their respective means due to this statistical artifact. The reason is that the extreme scores of the matched patients also include an extreme "score" on the "error component" or unreliable part of the measure representing the many unmeasured variables. By chance alone, this unreliable component will be less extreme the next time the measure is taken. This can cause or contribute to the finding of a statistically significant difference as the two groups regress to different means.

Figure C-1.—An Example of Statistical Regression Resulting From Matching



The report on the Duke registry by McNeer, et al. (240), clearly fits this picture. The surgical patients in that study were drawn from a "healthier" population. As Hammermeister, et al. (181), acknowledged, "there are probably additional unmeasured or undescribed variables of prognostic significance" in such data. Although these investigators are skeptical that this can alter the results, accumulated evidence indicates that regression can produce spurious statistical findings. There are no foolproof statistical remedies to this problem, but there are some recently developed analytic techniques that can partially adjust for measurement error (206). These approaches may be useful in situations where there are multiple measures of health status and a conceptual model specifying the presumed relationships among the variables involved. This technique would improve the statistical conclusion validity problems associated with this design.

The problems in matching indicate the difficulty in overcoming differences resulting from selection in a nonrandomized study design. The inability of statistical techniques to remove or adjust away these differences is graphically illustrated by the results of a recently reported study of the effects of drugs on coronary heart disease (73). Significant differences were found in the 5-year mortality rate for adherers (15 percent) and nonadherers (28 percent) in the placebo control group. A multivariate statistical analysis employing 40 baseline variables was performed to adjust for the differences in adherence. The adjusted mortality rates were only 16.4 and 25.8 percent, respectively. The baseline characteristics accounted for only a small amount of the initial difference. The authors noted that there must be unmeasured variables such as alcohol consumption and personality characteristics that can account for this difference.

Retrospective Case-Control Study.—Perhaps the most difficult variant of the cohort design is found in the field of epidemiology where retrospective case-control studies are frequently used to establish causal processes. This design consists of a group of people with a disease (i.e., the cases) who are compared with another group without the disease (i.e., the controls) to determine if they differ in their exposure to a presumed causal agent. The major problem in this design is in the selection of the comparison (or control) group. This is the major threat to the validity of this family of designs, because it is not possible to adjust for initial differences or to ensure that the treatment and control groups are equivalent. The problem faced by the epidemiologist-researcher is considerable, because the retrospective nature of the design implies no control of the treatment.

A recent dispute over the role of estrogen therapy for postmenopausal women as a cause of endometrial cancer illustrates the problems encountered in using this research design to assess technologies. The major point of contention among researchers (191,196) concerned the appropriateness of the control group. The traditional approach in this area had been to select women with other forms of gynecological cancer. Studies using this selection procedure have found a consistently high association between endometrial cancer and estrogen use.

This method of selecting controls has been criticized for not correcting a bias among the target cases that favors the obtained result. Specifically, it has been claimed that estrogen is associated with uterine bleeding and that this condition normally leads to careful overrepresenting in the population of confirmed endometrial cancer patients. To counteract this potential selection bias in choosing cases, Horwitz and Feinstein (191) recommend the use of women being treated for uterine diseases by either dilation and curettage or hysterectomy. These women, they argue, will include many referred because of vaginal bleeding. The use of such a population to create both treatment cases and controls will adjust for the bias resulting from increased surveillance and detection. Using both selection procedures, Horwitz and Feinstein demonstrated a reduction in the likelihood of estrogen causing cancer from about 11 to a factor of about 2.

Critics of this alternative selection approach claim that there is little or no detection bias since most cases of endometrial cancer are eventually diagnosed (196). They maintain that the alternative controls used by Horwitz and Feinstein are biased because they exhibit many benign conditions not normally detected. Moreover, estrogen may cause some of these other uterine diseases. Consequently, estrogen would be overrepresented in the controls. As Cole (61) has stated, patients undergoing the same diagnostic procedure as the cases can be "an inappropriate control group" since the same causal agent may be responsible for their illnesses.

In conclusion, it is important to note that the results generated by this design are essentially correlational and do not lead to unequivocal causal inferences. Horwitz and Feinstein[3] located 17 medical "topics" where multiple case-control studies reached differing conclusions. Selection bias (i.e., "avoidance of constrained controls") was the most frequent methodologic problem involved in the 17 disputes. The two approaches for constructing a control group discussed above can be viewed as providing a range of estimates for the relationship being examined. Because of the internal validity problems associated with this design, the use of different control groups to bracket the range of relative risk estimates should be considered. This would also improve construct validity in those instances where the effects of the technology are not well understood. Multiple case-control studies can also play a useful role in generating or confirming candidates (or potential causes) for unanticipated negative findings (e.g., toxic shock syndrome). In these instances, this epidemiologic approach is on the methodologic frontline of medical technology assessment. Often, where the event is rare and the number of cases is small, it is the only available method for making an assessment—e.g., of the role of aspirin in Reye's Syndrome. As with the Horwitz and Feinstein critiques, multiple studies using different controls were necessary before the association of aspirin to the disease was considered established.

Historical Controls.—Innovations often diffuse so rapidly and completely that the potential for untreated controls is greatly reduced or eliminated (see 253,266). In such a situation, researchers typically are forced to use a variant of the NECGD or cohort design that employs historical control groups—i.e., patients treated prior to the innovation. The important change in the design is a temporal one; patients in the comparison group are no longer treated concurrently with the experimental group. Some problems with the historical control group design are illustrated by a recent article discussing the use of adjuvant chemotherapy for treating osteogenic sarcoma (215).

Following the development of this treatment in the early 1970's, researchers began to experiment with ways to improve its apparent effectiveness. One approach was to treat patients with the drugs before their cancer had metastasized. Historical controls drawn from patient records dating from the 1960's were used in this research, and the results were provocative. Nearly half the patients treated lived 2 years without a recurrence of the disease, compared to only 20 percent of patients in 1960. Unfortunately, the change in therapy from 1960 to 1970 was also accompanied by other changes in diagnosis, treatment, and patients. The use of the computed angiographic tomography (CAT) scanner in the 1970's provided a much more sensitive test for detecting patients who did not have metastases. At the same time, surgeons began removing metastases in the lungs. At the Mayo Clinic, where both of these techniques were employed without chemotherapy, the survival rates equaled those of patients treated with the drugs. In addition, the patient mix probably changed over time so that those with the worst prognosis no longer constituted the majority of those treated. These criticisms of the research design and recent findings of a small controlled trial have convinced the National Cancer Institute to support a multicenter RCT to assess the efficacy of adjuvant chemotherapy for osteogenic sarcoma.

---

[3] R. I. Horwitz, and A. R. Feinstein, "Methodologic Standards and Contradictory Results in Case-Control Research," *Amer. J. Med.* 66:556, 1979.

This design demonstrates the importance of *history* as a plausible rival hypothesis in interpreting research results. It also points out that innovations in medical technology are not discrete events, but are often accompanied by other changes in the organization and delivery of medical practice that can affect construct validity. For example, surgeons note that there were major changes in the procedure for CABG surgery in the mid-1970's (e.g., cold-blood technique) and attribute to these changes responsibility for the decline in operative mortality. But, as this discussion has shown, the decline could also be due to a corresponding change in patient mix as more low-risk patients were convinced of the benefits of this innovation.

Wortman, Reichardt, and St. Pierre (402) have also suggested multiple measurements as a method of strengthening the basic NECGD. They recommend "double pretests" to estimate the change in baseline behavior of subjects in the absence of any treatment (by allowing each person to serve as his or her own "control"). The double pretest considerably strengthens the basic NECGD and should be employed whenever there is time to conduct two pretests prior to treatment. It is feasible when there is time to conduct two pretests prior to treatment. It is feasible when there is some lag between patient application and acceptance in a treatment program, as sometimes occurs in oversubscribed programs with long waiting lists. In situations where treatment is or must be made immediately available, the use of a double pretest would probably not be consistent with professional ethics.

Time-Series Design.—Often, data relevant to the assessment of a medical technology are collected at regular intervals over an extended period. Data archives such as the one used in the Manitoba evaluation of tonsillectomy can provide periodic information on the frequency and outcome of an innovation. If this is the case, a time-series design can be used. This design consists of multiple observations prior to and subsequent to the initiation of a treatment or other type of intervention (see table C-3 top row). Analysis of a time series involves checking for changes in either the level or slope of the series after the intervention.

Table C-2.—Relationship of Methods and Policy Issues to the Innovation Process

Level of development	Method/validity	Policy issue
New	Needs assessment Technical feasibility/ construct validity	Social need
Emerging	Research design/ internal validity Cost-benefit analysis Secondary analysis/ statistical conclusion validity	Efficacy, safety, social impact
Existing	Postmarketing surveillance Data synthesis/ external validity Needs "reassessment" Cost-effectiveness analysis	Effectiveness, safety

SOURCE: National Institutes of Health, "Criteria for Identification of Candidate Technologies for Consensus Development," memo, Feb. 23, 1978.

Using this design to study the impact of a hospital merger on a number of cost indicators, Whittaker (391) was able to demonstrate that, contrary to prior belief, the cost-per-stay increased after the merger as did total expenses per patient day. Employing this quasi-experimental design and the sophisticated statistical analysis procedures that have recently been developed for it, Whittaker demonstrated that a complex "organizational" innovation had uniformly "unfavorable" impacts.

The interrupted time-series design may at first seem to be an attractive assessment methodology that coincides with a number of convincing innovations—e.g., renal dialysis and the cardiac pacemaker represent successful medical technologies that appear to fit this design. However, upon reflection, it is clear that other information was available and used in the assessment of these innovations—i.e., physicians knew what happened to patients who did not receive the innovation—they invariably died. In such cases where the prognosis or time course of a disease is well documented, the technology evaluator has the benefit of a comparison series: a multiple time-series design (see table C-3). The comparison series helps to eliminate a number of threats to validity (e.g., history and maturation) and to reduce the plausibility of others. Time-series data can provide useful and inexpensive monitoring of an innovation and can even furnish evidence of causal effects. Thus, they could be used in the postmarketing surveillance of medical innovations.

Statistical analysis of time-series data is still a rarity in the assessment of medical technology. Although

Table C-3.—Multiple Time-Series Design

0	0	0	0	0	X	0	0	0	0	0
0	0	0	0	0		0	0	0	0	0

SOURCE: Adapted from T. D. Cook, and D. T. Campbell, *Quasi-Experimentation: Design and Analysis of Research in Field Settings*, 1979.

there have been a few exceptions such as Albritton's[4] study of the 1966 Federal program for measles immunization, most researchers have been content to present their data in graphic form (see 342). The effects of interventions are often dramatic, and visual judgments of statistical and medical significance may be adequate in many cases. However, statisticians (see 330) have long warned that graphic representations of data can often be misleading. This warning has been specifically repeated with respect to time-series analysis (175,202). These authors demonstrate that visual and statistical analysis of time-series data often lead to opposite conclusions. More detailed descriptions of interrupted time-series analysis and examples of applications may be found in Glass, Willson, and Gottman (167), and McCleary and Hay (237). Recent tutorial articles such as the evaluation of a Regionalized Perinatal Care program in North Carolina (161) provide examples of the growing use of this design in assessing medical interventions.

Conclusions.—Although the cohort or NECGD is often easier to use than an RCT, it suffers several weaknesses in the form of threats to validity. The most serious threat are *selection* differences. Because subjects are not randomly assigned to treatment and comparison conditions, pretest or baseline differences among the groups are quite likely. These initial differences are then confounded with changes due to treatment observed at the posttest. A number of analytic approaches have been suggested to deal with this problem. For example, the Cox regression technique has been used to analyze the survival data from cohort studies (see 181). However, the analysis rests on the assumption of proportional hazards, that both groups have the same risk of illness, and this is unlikely to be true where the groups are nonequivalent. The results from such nonrandomized experiments thus remain extremely equivocal, particularly when the experimental and comparison subjects differ significantly in terms of important pretreatment characteristics.

Because there is no agreed upon analytical solution to the problem of baseline selection differences, probably the best that researchers can do currently is to use several different methods of analysis (70). If the results from the various methods are congruent, evaluators may state their conclusions with appropriate caution. If different methods lead to different results, the situation is more confusing, and the technology assessment will have to be more tentative in its conclusions. Although this design may be appealing, it poses such severe problems in analysis (i.e., it has doubtful statistical conclusion validity) that extreme care is warranted (402).

In sum, do nonequivalent controls, particularly with matched groups, provide useful information for a technology assessment? Our general answer is that they do not. The methodological problems resulting from these designs are often of such serious concern as to undermine the credibility of the findings. Only when the competing explanations or rival hypotheses (i.e., important threats to validity) can be demonstrated to be implausible or can be ruled out through other subsidiary data should such studies be considered seriously in a technology assessment. Statistical solutions, in particular, should be viewed with skepticism despite the impressive impenetrability of their algebra.

As this discussion has shown, there is justifiable concern about the credibility of the evidence produced by controlled nonrandomized studies. For many experts and informed practitioners, the potential existence of such methodological problems is sufficient to cast doubt on the findings. These concerns in determining the efficacy of medical innovations are not new. Meier's (243) discussion of the Salk polio vaccine trial of 1954 indicates that the original quasi-experimental design was upgraded to an RCT in certain States because of these concerns. The original design called for second-grade children to receive the vaccine with first- and third-graders as comparison groups. The difference in the size of the effect observed from these two designs illustrates the problem with estimates of efficacy obtained from cohort quasi-experiments. The differences in the incidence of polio cases was 40 (per 100,000) for the RCT, while it was only 27 for the alternating grade cohort quasi-experiment. The nonrandomized results thus underestimated the efficacy of the vaccine by nearly 50 percent.

The multiple time-series quasi-experiment is much stronger on internal validity than the cohort design. The research by Sherman (342) and his associates (343) demonstrates the potential utility of the time-series design in the evaluation of health programs at the individual patient or program level of analysis. The time-series design is relatively unobtrusive; it rarely requires the changes in operating policy that a randomized experiment often does. In addition, time-series analyses may be used to assess innovations that have already been in operation for a considerable length of time (see 403). Since many observations are required to perform time-series analysis, it is most appropriate for agencies that collect data at regular intervals. Hospital and insurance reimbursement or claims records would be most useful. These could be used to provide information on cost, utilization, and health outcome.

UNCONTROLLED DESIGNS

The most common form of evaluative study for medical technology assessments employs a nonexperimental or uncontrolled design (401). These so-called case studies do not include any comparison groups at all and usually report judgments by physicians about the extent to which each patient improved.

---

[4] R. B. Albritton, "Cost-Benefits of Measles Eradication: Effects of a Federal Intervention," *Policy Analysis* 4:1, 1978.

The first studies on gastric freezing (143) were almost all of this type. The results of the early studies were largely on the self-reports of a few patients subjected to the procedure. The basic case study may be described as a "posttest only" design, since only one measurement of the status of the subjects is used. A slight elaboration of the "posttest only" case study is the one-group pretest-posttest design, which includes a measure of the status of the patients prior to, as well as after, treatment. The use of two assessments allows the researcher to estimate changes in the patients over the course of treatment, as well as their final status.

The problem in using nonexperimental evidence to assess a medical innovation is illustrated by a new technology to facilitate the management of diabetes—home blood glucose monitoring. This is a fairly recent innovation that shows promise of helping diabetics monitor their blood glucose levels more accurately than before, thereby allowing them to participate in their treatment by changes in diet and exercise (362). To obtain information on blood glucose level, the patient pricks a finger and applies the blood to a reagent or chemstrip. Glucose level can then be determined either directly or by reading a reflectance meter. This technique is viewed as a replacement for urine testing, although it costs three to four times as much.

There is a great deal of enthusiasm about blood glucose monitoring, and it is being introduced in a number of diabetes outpatient clinics. However, there is little evidence as to its effectiveness. Only nine studies of this innovation could be found, and none of them used a control group. Furthermore, many of the studies simultaneously introduced other regimens with the blood glucose monitoring procedure, thereby raising construct validity questions. The other regimens introduced included exercise, group therapy, and spray injection of insulin. Any of these techniques could have produced the beneficial effects reported. Moreover, most of the studies had very few patients; five had 17 or fewer subjects. Thus, selection of highly motivated patients, for example, could produce overly optimistic results.

Although nonexperimental studies can provide information concerning the technical feasibility of a new medical technology, they are far from definitive. They can also provide useful "qualitative" information (286) concerning the acceptability of the technology to patients (e.g., their willingness to draw repeated blood samples from their finger), factors affecting compliance (e.g., the interpretability of the chemstrip), and related behavioral issues that may hamper its utility. However, these studies should not be viewed as providing adequate information concerning efficacy and safety. Without a valid comparison group, it is not possible to determine whether the benefits are due to patient self-selection or to other factors. Nor is it possible to tell whether the innovation is superior to the urine testing methods now commonly used.

The major difficulty with nonexperimental designs is that they are subject to practically all of the threats to internal validity described above. It is inappropriate to interpret such studies as indicating that observed changes in patients are due to the innovation. Unfortunately, such interpretation is a common occurrence. Physicians, lacking training in research methods, can mistakenly perceive such preliminary pilot studies as being definitive. This can result in premature diffusion. The situation is often exacerbated by the exaggerated claims made by the developers of the technology.

Research Synthesis: Integrating the Results of Multiple Studies

The preceding section emphasized a set of principles underlying the design and interpretation of individual studies to assess the effects of medical technologies. The assessment of medical technology, however, is a process that involves more than the consideration of a specific research study (see 266,269). In order to conduct a technology assessment, multiple sets of evidence, where available and relevant, must be considered and synthesized. Although little attention has been given to methods for synthesizing evidence about the effects of technology (266,398), there do now exist formal techniques to integrate the findings from different studies and to develop generalizations based on their results. The methodological and conceptual issues involved in the conduct of such analyses are considered in the discussion below as part of the technology assessment process.

The synthesis of research data is often both controversial and complex. Controversy arises because the results of studies about a particular technology may vary and/or be interpreted differently by different assessors. Synthesis is complex because medical technologies may have different clinical outcomes depending on who uses them or when they are used. Establishing the efficacy and safety of a technology on the basis of research evidence is typically a lengthy process. These assessments (i.e., safety, etc.) depend basically on the amount and quality of the research evidence and the analyst's ability to deal with the available information (i.e., ability to determine the validity of the evidence and to combine the various types of information appropriately).

This section focuses on problems of synthesizing independent research studies relevant to a technology assessment and describes some formal procedures that enable systematic integration of research results. In addition to describing quantitative methods, it considers a number of methods for synthesizing information that rely on group decisionmaking approaches to technology assessment. These methods are often used to resolve controversies about research evidence and to develop guidelines for employing particular medical technologies.

Although information from a variety of sources must be considered as part of a technology assessment (e.g., costs, social impact, etc.), research data concerning efficacy and safety form the central component. These outcomes are essential for determining social impact (see 295). Methods for synthesizing research information and using it in decisionmaking are critical to the outcome of an assessment. One purpose of the following discussion is to suggest what types of data are useful in the development of technology assessments and how they should be treated. The section is organized in four parts: 1) the application of the validity concepts introduced in the previous section to the synthesis of multiple research studies; 2) current approaches and problems to synthesizing research evidence; 3) quantitative research synthesis and integration methods; and 4) formal, group decisionmaking methods for synthesizing research evidence.

Validity

The previous section of this appendix emphasized inference problems inherent in the interpretation of individual studies of medical technology. From a methodological perspective, true experiments, particularly RCTs, reduce problems of equivocality of inference, as compared to other research strategies. A single RCT, however, cannot resolve all questions about a technology, and technology assessments cannot rely solely on their availability. If randomized studies are not available, decisions will have to be made about how to treat the validity problems inherent in other types of research. For example, the reduction or elimination of threats to internal validity by an RCT does not automatically avoid problems due to low external, construct, or statistical conclusion validity. In particular, external validity often must be established by examining evidence from multiple studies. Thus, validity considerations are as relevant to the problems of aggregating and synthesizing the results of many studies as they are to interpreting a single study. In the following discussion, the validity framework is extended to indicate its use in integrating evidence from multiple studies.

INTERNAL VALIDITY

Internal validity problems are central to the synthesis of findings from multiple studies. Because of the limited availability of RCTs, other evidence that can reduce the number of plausible alternative explanations for findings should be considered. However, the validity of nonrandomized studies must be carefully examined. If all, or most of the evidence about a particular technology was generated through similar, and perhaps consistently flawed research designs, the advantage of multiple sets of data may be lost. If the available literature includes a large number of studies with low internal validity, then a simple aggregation of the results may yield a conclusion open to a variety of alternative interpretations, especially when similar validity problems affect each study.

The existence of a few studies using randomized control group designs, on the other hand, does not guarantee high internal validity. Again, the alternative explanations must be considered to determine if they can be eliminated. For example, the previous section noted that the RCTs assessing the efficacy of CABG surgery were consistently flawed by differential patient attrition (or experimental mortality). Only through the availability of other evidence provided by additional control groups or improved analyses can the remaining threats to internal validity be eliminated (see 71). The principal problem in data aggregation is to identify such validity problems and to develop a strategy for aggregating the results of studies that differ in their internal validity.

In most cases, it is likely that experimental, quasi-experimental, and nonexperimental data will be available. The problem, then, is the appropriate choice of both the evidence and the amount of emphasis it should be given. One possibility (166) is to aggregate studies that are high in internal validity separately from more "poorly controlled" ones.

Gilbert, McPeek, and Mosteller (159,160) provide evidence of the importance of this strategy for medical technology assessment. These investigators compared the results of randomized and nonrandomized clinical trials of a series of medical innovations. They found that positive results were more likely to be obtained by an uncontrolled research study than by an RCT. RCTs tended to yield much less favorable conclusions about effectiveness. For example, among 53 studies of portacaval shunts, they found only 6 well-controlled trials. Of these controlled trials, three were associated with negative conclusions about the treatment and three yielded moderately positive conclusions. This compares with 32 uncontrolled studies, where 24 were very positive, 7 were moderately positive, and 1 was negative. In general, Gilbert and colleagues found that the poorer the methodological quality (i.e., the lower internal validity), the more likely that a treatment would appear to be effective. The implication is that conflicting claims surrounding medical innovations may merely reflect differences in the validity of the research designs.

STATISTICAL CONCLUSION VALIDITY

In evaluating a set of studies, it is necessary to consider whether serious threats to statistical conclusion validity exist in individual studies and whether these threats prevent developing conclusions about the technology under study. Berk and Chalmers (26) examined the adequacy of the statistical analyses in a research synthesis of studies dealing with the cost effectiveness of ambulatory care (see below). They reviewed those studies reporting no difference in clinical outcome to determine whether there was sufficient statistical power to detect a 25-percent difference (if one existed). Of the 23 randomized trials, 16 had sufficient power. Seven RCTs plus all the nonrandomized controlled trials were classified as having "indeterminant clinical outcomes since selection bias may influence the outcome and obviate statistically valid comparisons when controls are not selected at random."

EXTERNAL VALIDITY

External validity is essential in assessing data from multiple studies. The more widely a treatment has been tested, the easier it should be to establish the degree to which results are generalizable to various populations and settings. Studies high on internal validity, such as RCTs, may often yield differing and apparently conflicting results, because different patients, settings, or procedures are used. It is crucial that these studies be aggregated or stratified according to external validity factors. The differences can often be dramatic. The NIH consensus conference on CABG surgery (96) found the surgery effective for patients with left-main coronary artery disease, but not for patients with single- or double-vessel disease. Similarly, radical mastectomy, once the universally recommended procedure for breast cancer, is no longer endorsed by experts (96) for women whose disease is detected early (i.e., Stage I and II).

CONSTRUCT VALIDITY

Construct validity is also a serious concern in synthesizing the results from many studies. As Pillemer and Light (292) have noted, it may be that differences across studies are due to the use of treatments that have only been labeled similarly. For example, surgeons at the consensus conference on CABG surgery, noted above, dismissed most of the studies raising concerns about the procedure's safety (i.e., high operative mortality), because the studies were conducted before a major change, called the "cold-blood technique," was adopted in the mid-1970's. The modification of a new technology can greatly affect its performance. These changes, which are often unreported, mean that an evaluation is, in fact, assessing a family of technologies, or a "moving target" (380). The rapid development that characterizes the early stages of technological innovation can lead to errors in data aggregation, because unreported, new components of the treatment may have been incorporated into various assessments. Often these labeling problems are more insidious in that other unnoticed technological changes co-occur with the innovation (e.g., improved diagnosis in osteogenic sarcoma).

Outcome Measures.—The major problem confronting those desiring more systematic methods for synthesizing the results from many research studies has been the inability to combine many different measures of efficacy and safety. One must ensure that comparable measures of the appropriate construct have been employed. For example, Berk and Chalmers (26) report a systematic review of the efficacy of ambulatory care as a cost containment measure to reduce inpatient expenditures. They found 134 relevant articles. Studies lacking either construct or statistical conclusion validity were eliminated. Of the 109 actual studies reported, 31 were eliminated because economic outcomes were not discussed. In the remaining 78 investigations, they found an appropriate measure of costs in only four studies! Thus, the improper measurement of a construct can significantly reduce the validity and usefulness of many studies for synthesis.

Problems With Traditional Synthesis Procedures

The traditional approach to synthesis is the literature review. Almost all technology assessments begin with such a research summary. Unfortunately, these reviews tend to be asystematic and subjective. Reviewers select the evidence they believe to be most relevant and typically organize their presentation around the demonstration of a particular hypothesis. Although technology assessments, such as those developed by the former NCHCT, were based on summaries assembled and reviewed by several experts, there are still a number of problems in relying on this approach to derive the implications of research and to resolve controversies.

METHODOLOGY

A central problem in literature reviews is how to deal with methodological issues. As noted above, the relatively few RCTs generally available (401) present an important obstacle to the reviewer. Well-controlled research studies are probably the best way to produce unequivocal evidence. However, the weight of other evidence may sometimes hinder their use.

This problem is illustrated by the controversy over electronic fetal monitoring (EFM). Several recent reviews of the efficacy and cost effectiveness of EFM (e.g., 17,367) have indicated that significant risks are associated with monitoring (in particular, an increase in cesarean section rate) and that it is not a beneficial diagnostic tool for many of the patients with whom it is being used. Significantly, reviewers who are skeptical of the use of EFM are primarily researchers; reviewers who are clinicians have come to a different conclusion and have strongly supported the broad use of EFM (see, e.g., 189). Researchers appear to disregard much of the published literature because it consists of reports of uncontrolled research. Wortman (401) notes that 23 of the 24 poorly controlled studies supporting EFM—there were no well-controlled studies supporting it—employed historical controls. From the perspective of a research methodologist, the lack of internal validity indicates that there is no valid basis for comparing monitored to unmonitored births. Clinicians, in contrast, appear to be swayed by the large number of case studies that describe successful applications/assessments of EFM. Since the rate of false positives leading to cesarean section is relatively low, this literature probably is most consistent with their own experience.

Even when RCTs are available and the weight of the evidence is not as discrepant as in the EFM situation, they may not fully answer questions about the technology. Tonsillectomy is a case in point. A substantial literature exists about the safety and efficacy of tonsillectomies, and experimental, quasi-experimental, and nonexperimental research is available. Cochrane (60) reports three different clinical trials on tonsillectomies conducted in England during the 1960's, but he contends that none of the trials resolved the policy controversy over the appropriate use of tonsillectomy. According to Cochrane, the available RCTs exhibit two methodological problems: 1) the treatment was compared with no or inadequate medical treatment (instead of an alternate treatment); and 2) the patients' parents were not blind to the conditions of the experiment, so those whose children were on the waiting list may have exaggerated their children's symptoms.

Wennberg, Bunker, and Barnes (389) note that a large-scale clinical trial is currently being conducted, but that the trial, in itself, will not resolve the controversy. This is because the current RCT does not include a sample of the full population of children for whom tonsillectomy is recommended. In essence, several internal and external validity problems prevent these available and pending RCTs from being unequivocal tests.

TIMELINESS

Some (e.g., 34) believe that clinicians, over time, will be able to determine which medical treatments are useful and which are not. The implication is that methodological considerations are not central. Others (e.g., 389) have suggested that this approach is ineffective and that many common medical practices are inadequately evaluated and perhaps worthless or unsafe. A question exists as to whether systematic reviews of research evidence can influence medical practice.

Several studies have examined the use of research by clinicians. Fineberg, Gabel, and Sosman (145), for example, reviewed the use of scientific papers by anesthesiologists. They found that there is a significant lag between research discoveries and their publication. Their view is that scientific papers affect actual practice slowly. This would seem especially true of reviews that attempt not only to summarize but also to draw implications from the literature. In part, this is because it takes considerable time for a published literature on any medical technology to develop. In several now-classic cases (e.g., gastric freezing), literature reviews were only published years after a procedure was abandoned because it was ineffective or unsafe (see 143, 245).

In the gastric freezing example noted above, either a more timely RCT (i.e., earlier) and/or more systematic attention to the available nonexperimental data might have hastened the abandonment of the procedure. These two evaluative processes are, in fact, related. Thus, if an RCT is not conducted during the initial investigational stage of a developing technology, then it is even more important that systematic attention be given to whatever data are generated, since these data may indicate whether or not an RCT is needed. It should also be clear that an RCT may not "solve" the technology problem, and, in many cases, research synthesis may stimulate the generation of additional data that are needed.

CONCLUSIONS

The evaluation of safety and efficacy evidence to understand the effects of particular medical technologies is complex. Complexity is related to the presence of bias and methodological problems, such as the lack of appropriate control groups in research reports and literature reviews. Research evidence may exist for any medical technology, but it may be difficult or impossible to synthesize these data without carefully considering the validity of the individual studies. Developing research that can be used for a technology assessment is obviously difficult, but is only the first step. Synthesis strategies are clearly necessary as part of this process to deal effectively with the results of the many studies bearing on a technology.

Systematic Procedures for Data Synthesis

A major implication of the previous discussion is that the need for policy-relevant information often outstrips the capabilities to provide it. The development of conclusive evidence about a technology at present seems to be a relatively slow process. This discovery process is probably better geared to the occurrence of "breakthroughs," those rare single studies or programs of research that resolve a controversy, than to dealing with elaborate arrays of potentially conflicting or inconsistent information. Procedures are needed that enable the accumulated insight gained from research to be usable within the technology assessment process.

The problems and benefits in systematically organizing and integrating research findings are discussed below. The procedures described, although not representing a panacea for all the problems identified, suggest how the process of research synthesis can be more rigorous. Some elementary qualitative procedures, as well as sophisticated statistical techniques, for conducting research synthesis are described below. The goal is to outline the range of systematic methods that may be employed and to contrast them with more traditional techniques.

VOTING METHOD

A simple form of synthesis has been called the voting method (226). This technique essentially involves organizing a body of literature according to some prespecified set of criteria. Usually, vote counting involves the selection of a particular sample of outcome studies, coding some aspects of their design and/or conceptual framework, and classifying the observed outcome(s) according to whether they are favorable, neutral, or unfavorable (i.e., "taking a vote"). The Gilbert, McPeek, and Mosteller (160) study, referred to above, is an example of this type of synthesis. Sampling the literature to determine the rate of successful innovation in anesthesia and surgery, their analysis indicated that about half the innovations assessed by RCTs were successful when compared to a "standard" treatment.

A frequent use of the voting method is to demonstrate differences obtained by various methodological approaches. For instance, Gifford and Feinstein (157) critiqued studies of anticoagulant therapy for acute myocardial infarction (MI). They examined all available literature on acute MI that reported control group studies of acute MI treatments. For each of 32 studies located, they coded the degree to which the diagnostic criteria for MI were clear, whether randomized control groups or other methodological criteria were employed, and summarized their findings in several contingency tables. The results of the vote count indicated that anticoagulant therapy was superior to no treatment more often in reports that did not observe methodological standards than in those that did.

The strength of vote-counting analyses lies in: 1) the precise identification of the populations of studies to be sampled, and 2) the coding of substantive and methodological aspects of the study according to clearly defined procedures. More widespread use of the technique could probably aid in determining which specific patient populations and/or conditions could be effectively treated by a medical technology. The voting method helps to avoid the problems of reviews that only selectively describe research or pay attention only to some aspects of the study. In addition, such analyses may be particularly useful in identifying relationships between methods and outcomes.

Krol (216) cites three problems with the voting method: sample size, effect size, and Simpson's paradox. Large studies are likely to produce statistically more significant results than those with small numbers of subjects due to differences in statistical power. Thus, a finding of no difference among treatment and control conditions will be correlated with small sample size. In fact, Hedges and Olkin (186) have demonstrated that the voting method itself generally lacks statistical power. A second problem is the all-or-none nature of the method. Some findings may show small, marginal effects and others large ones, but they would count the same. Consider the case where effect size is correlated with outcome—large, positive effects and small, negative ones. The voting method would yield no difference when, in fact, there was an overall positive effect. Simpson's paradox is a more subtle statistical point in which it is possible, under certain conditions, to reach different conclusions by aggregating data from each study rather than by counting each study separately. The paradox results from unbalanced cell frequencies. Finally, Light and Smith (226) have noted these and some additional problems with the method. The most important is that vote counting may oversimplify the results of studies and cause one to overlook more subtle, but important, relationships (especially interactions among variables).

META-ANALYSIS

A second synthesis technique, called "meta-analysis," has been developed by Glass (164,165). Meta-analysis or the "analysis of analyses" is a rigorous statistical approach to research synthesis. Meta-analysis utilizes the actual results of studies and permits the determination, across a set of studies, of the magnitude-of-treatment impact. Most statistical analyses, as summarized in research reports, ignore both the size and direction of effects and yield only a global probability of a "significant" difference. Meta-analyses are useful for assessing treatments where a large number of studies exist and where findings across studies seem to have great variability. As used by Glass, such analyses require that comparison groups be available (i.e., either randomized or quasi-experimental groups) and that the original research reports contain appropriate statistical information such as the group means and standard deviations. Glass (164) describes some indirect procedures for deriving the effect size from the inferential statistics reported in a study (i.e., t-test, F, etc.).

Effect sizes (ES) are calculated by determining the difference between the mean of the treatment group ($\overline{T}$) and the mean of the comparison group ($\overline{C}$), divided by the standard deviation of the comparison group (SD_c). Thus,

$$\frac{\overline{T} - \overline{C}}{SD_c}$$

This procedure converts the average effect of each outcome measure into a common scale (i.e., standard deviations) that can be compared to results of other studies. If a treatment has no effect, then there would be a zero effect size; if the treatment is effective (i.e., better than the current alternative), the effect size is positive; and, if the treatment is inefficacious, the effect size is negative. By making some assumptions about the skewness of experimental and control group scores within each study, and the distribution of effect sizes across a large number of studies (i.e., that they are normally distributed), effect sizes can be converted into percentile ranks and inferences can be made about the overall effects of a medical technology.

One of the best recent health technology examples of a meta-analysis is Smith, Glass, and Miller's review (354) of the outcome studies of psychotherapy treatments (see also, 353). Smith and colleagues searched the published literature, including abstracts, and included within their analysis all available control group studies of the effectiveness of any form of psychotherapy. Drug studies were analyzed separately, while those studies that did not involve the use of professional therapists (operationally defined as psychologists, psychiatrists, and social workers) were eliminated from the analysis. The investigators coded an extensive number of variables for each study, including methodological criteria such as the nature of the patient assignment to condition (e.g., random v. matching), experimental mortality, and other threats to internal validity. Effect size scores were calculated for each principal dependent measure. The analysts also developed a code for validity of the outcome measures.

Smith, Glass, and Miller's (354) findings indicated that, on the average, the difference between scores of the groups receiving psychotherapy and scores of the control groups was 0.85 standard deviation units. Assuming the normal distribution of effect size scores, this average standard score indicates that a typical person who receives psychotherapy is better off than 80 percent of the people who do not. Smith and colleagues also conducted a number of analyses to determine whether the methodology of the study affected results and whether different therapies (or other factors) were differentially efficacious. They found few reliable methodological differences. It appeared that outcomes were not related to the use of randomized control groups. This finding should, however, be tempered by the knowledge that all of their sample studies used comparison groups and were generally high in internal validity. When this is not the case (i.e., where quasi-experiments are included), then the outcome can vary with the methodology (i.e., research design). Wortman and Yeaton[5] have shown this to be the case for the studies on CABG.

There has been some criticism of Smith and Glass' (353) approach based on their "lumping together" of a large number of what some consider incomparable treatments and outcomes (e.g., see 137). The strength of the effect size technique, however, is that it provides a common metric that permits analysis of the differences (methodological and substantive). Smith and colleagues' classification variables for each study were fairly comprehensive and yielded a systematic comparison of studies on the basis of their conceptual and methodological designs. What is problematic about such meta-analysis, however, is that the findings are heavily dependent on a number of decisions that are not always made explicit. These include the studies selected/rejected from the literature, variables included/excluded, and their construct validity. It is not possible to ascertain biases resulting from Smith and colleagues' sampling decision nor whether only certain types of studies, therapies, or variables are assessed using control group designs (273). A broader analysis of psychotherapy research might yield different conclusions than those drawn by these investigators.

---

[5] P. M. Wortman, and W. H. Yeaton, "Synthesis of Results in Controlled Trials of Coronary Artery Bypass Graft Surgery" (Ann Arbor, Mich: Institute for Social Research, 1982) (report submitted for publication).

OTHER SYNTHESIS TECHNIQUES

A number of other methods exist for statistically combining the results of independent studies (see 69,292,324). The effect size method described above actually incorporates several procedures. The most important of these methods is the comparison of treatments to detect interactions between characteristics of a study and outcome (i.e., external validity issues). As noted in the earlier discussion of the voting method, some of these procedures can be employed when effect scores are not computed. Additional statistical methods combine probability values from various studies and adjust outcome scores according to the relevance of the data.

Rosenthal (324) describes a number of procedures for combining probabilities. These range from adding observed probability (p) levels across different studies to adding weighted standardized (z) scores. These methods also include the testing of mean probability values. Essentially, using such procedures allows one to indicate whether significant effects are obtained across a set of studies. The problem in using probability values is one of statistical conclusion validity. The number of subjects per study influences the statistical power to detect whether significant overall differences are present.

DuMouchel and Harris (131) discuss another interesting quantitative method for synthesizing the results of experiments done with human and animal species. This method, a sophisticated application of Bayes' theorem, provides estimates of carcinogenic risk from various substances derived from the results of epidemiological studies.

IMPORTANCE OF SYSTEMATIC DATA SYNTHESIS

Earlier, it was noted that technology assessment is essentially a synthesis process that involves the review and integration of research findings. There are a number of specific benefits that result from employing formal procedures for data synthesis (see 292). The first advantage is that formal syntheses help to identify contradictions in the literature by systematically organizing studies according to specified classification factors. It becomes possible to segregate differential outcomes according to treatment characteristics and/or methodological approaches. The analysis of different findings when controlled and uncontrolled studies are employed (see 160,400) is a good example of this aspect of meta-analysis.

A second benefit of meta-analysis has to do with the use of effect size scores. Not only do such scores provide insight as to the worth of the treatment, as in the Smith, Glass, and Miller (354) psychotherapy example, but they also provide a benchmark for later research. Thus, for example, a meta-analysis conducted by Posavac (294) of 23 controlled studies of patient education programs found a 0.75 average effect size. Posavac indicates that this should provide a standard against which new patient education programs can be assessed. If the effect sizes of new programs are only 0.20 (and similar dependent measures are employed), this would probably indicate that the programs are not particularly effective, at least for the problem or population for whom they were designed.

Another advantage of quantitative synthesis methods are that they serve to control for certain statistical conclusion validity problems (e.g., power) that some commentators have reported as severe in the medical literature (e.g., 141,172,337). It can be assumed that the widespread use of meta-analysis and other quantitative approaches to synthesis would improve statistical reporting practices by calling attention to different investigators' use of data. In addition, errors in analyses, such as the use of multiple independent inferential tests without appropriate error rate control or incorrect inferences because of a lack of power, would be compensated for by most meta-analytic procedures. Although errors in data collection and, perhaps, in computation of means and standard deviations would not be corrected by these synthesis methods, the systematic analysis of multiple studies should render the effect of such errors less consequential. The attention to systematic considerations of the "weight" of evidence across research studies should have a general salutary effect.

Finally, it should be noted that, although these procedures seem most appropriate for evaluating more mature technologies that have accumulated a considerable body of research, they are often applicable to less developed technologies. In some cases, where only meager evidence is available from a small set of studies, it may be that a review of specific components from some other portion of the literature may suggest the effectiveness of the new technology. Thus, physiological evidence may be considered with other clinical, experimental data as in the case of radical mastectomy (see 147) noted earlier.

Group Decision Methods

Although the application of formal statistical procedures for the integration of data from individual studies should improve the ability to conduct technology assessments, the use of such methods does not entirely resolve policy controversies. Such analyses cannot go beyond the available data on a particular problem, nor can they substitute for informed judgment. In the discussion below, some recently suggested procedures for resolving conflicts across research studies and for developing assessments of particular technologies are described. These informal methods include a new approach to decisionmaking sponsored by NIH, referred to as consensus development, and a number of other decisionmaking techniques (e.g., Delphi) that have been employed in assessments of medical technology.

NIH CONSENSUS DEVELOPMENT

In response to congressional pressure to assist in the transfer of technology, NIH initiated its consensus development program in 1977 (310). Perry and Kalberer (287) recently described the consensus development program at NIH. Its goal is to bring together various concerned parties (e.g., physicians, consumers, bioethicists) in order to seek agreement or "consensus" on the safety, efficacy, and appropriate conditions for use of various medical procedures. Judgments about the technology under consideration are intended to be based on the scientific evidence of its effectiveness as well as on information about its social, ethical, economic, and legal impacts. The consensus development process is designed to produce a written recommendation, called a "consensus statement," that can be accepted by clinicians and researchers. The statement is supposed to identify both what is known and not known about the technology.

Topics for NIH consensus development are chosen because of their current or potential importance (e.g., in terms of cost, number of patients affected). Since September 1977, NIH has held more than 30 consensus conferences at which the evidence and implications of a wide variety of technologies have been considered. Topics have ranged from bee sting kits to CABG surgery. The technologies include both emerging, as well as currently used, technologies that either have not been carefully evaluated for safety and efficacy or are controversial. Recently, there has been a trend toward more mature technologies (see next major section below) for which there is more scientific evidence concerning effectiveness.

Over the past few years, the conferences have generally followed a similar format. A panel of neutral experts is selected by NIH to hear presentations by the leading medical researchers addressing a prespecified set of questions about the technology. The presentations, usually summarizing the latest research findings, are made over a 2-day period during which both panelists and audience members discuss the research findings. On the evening of the second day, the panel is sequestered to draft a statement responding to the questions. Usually, they deliberate through the night, writing as many as four drafts of the consensus statement. In some rare cases, minority reports are developed to indicate disagreement with the majority recommendations. The next morning the statement is read to the audience for their comments and criticisms. The conference concludes with a press conference. The panel then disperses with the final task of revising the statement. The consensus statements are widely disseminated by NIH through direct mail to thousands of organizations and individuals and by publication in leading medical journals such as the *New England Journal of Medicine* and the *Journal of the American Medical Association*.

From a methodological perspective, two aspects of the consensus development process are of concern: 1) its sensitivity to the limitations of the research evidence, and 2) the extent to which a comprehensive and systematic review of the research literature is considered. There is little published evidence concerning these issues. An examination of panelists participating in previous consensus conferences (96,97,98) indicates that there has been no consistent policy to include a methodologist—either a biostatistician or epidemiologist. On few panels were such persons included. This means that in most cases there was no informed person who could indicate the methodological limitations of a study. The problems to which methodological ignorance can lead have already been described.

The consensus conference on CABG surgery was an exception (95). Two biostatisticians are listed as members of the panel, and their influence on the consensus statement is evident. The methodological limitations of the research literature with respect to a key question are discussed at length (see 95). A number of these methodological problems have been noted above: attrition due to crossovers, use of historical controls, statistical analyses of registry (i.e., quasi-experimental) studies, and the like.

Despite this indication of methodological detail, there is apparently no formal policy to provide systematic reviews of the research literature. For example, the weight of the evidence on the efficacy of the coronary bypass procedure as presented in the published consensus statement (95) was evidently derived from two large, multicenter RCTs: the somewhat controversial VA study (255) and the ongoing European trial (136). Our examination of the literature revealed that there are at least 30 studies, of which 9 are RCTs. Given the emphasis on external validity issues (i.e., identifying the patients for whom the surgery is beneficial), the limitation of the discussion to two studies was clearly unwarranted.

This problem has occurred in other consensus conferences as well. In a recent letter to the *Journal of the American Association,* Jones (203) noted that one of the conclusions from the conference on adjuvant chemotherapy of breast cancer was "based on incomplete information." He pointed out that the results of only five studies were presented while there were "at least nine major studies" containing "convincing evidence" on the effectiveness of chemotherapy in postmenopausal women. (The consensus statement claimed effectiveness for only "a select group of breast cancer patients.")

On the other hand, in most consensus conferences, the attention to these methodological concerns has been reversed. Most consensus statements reveal little discussion of methodological issues and limitations of the studies even where this might be appropriate. However, extensive background materials are often made available to the panel. These included a computerized bibliography of the literature and reprints of the articles.

The consensus conferences are coordinated by NIH's Office for Medical Applications of Research (OMAR). Although the topics are selected by the relevant institutes, OMAR makes the final decision about the suitability of the topic, panel composition, and the proposed format for a consensus conference. Over the past 2 years under OMAR's direction, the conferences have developed in a number of ways. The use of a fixed format has already been noted. Other approaches involving adversary (i.e., nonneutral) panels and task forces have been almost entirely abandoned. Moreover, the questions that have been posed to the conferences have been addressed strictly to those issues on which there is enough factual evidence to reach agreement. This has resulted in the omission of controversial issues. For example, in the recently published statement from the Reye's Syndrome consensus conference (67) questions about the role of salicylates (i.e., aspirin) were deliberately omitted because OMAR felt little was known about it (although the limitations of the studies establishing this association were briefly discussed). An editorial on the coronary bypass consensus statement in the *New England Journal of Medicine* (308) complained that it and other consensus statements "represent the lowest common denominator of a debate—the only points on which the experts can wholeheartedly agree." This reflects the current orientation of OMAR away from "state-of-the-art" conferences. One methodological consequence is that gaps in knowledge and needs for further research may not be as readily identified.

FORMAL GROUP DECISION METHODS

In addition to the NIH consensus development process, a number of systematic procedures for developing consensus based on behavioral science principles (see, e.g., 163) have been developed. The goal of these procedures is to aid groups composed of individuals with different information and perspectives to develop group judgments that best take account of the positions of the individual members. In the discussion below, two methods—Delphi and nominal group technique (NGT)—are presented. These techniques illustrate the potential and limitations of these methods for technology assessment.

Delphi Technique.—Delphi (78) is probably the oldest structured model for involving groups in decisionmaking processes and has been used widely in health care. The Delphi technique uses a series of questionnaries (or individual interviews), each followed by anonymous feedback summarizing all the participants' responses. Although Delphi was originally developed by the Rand Corp. to synthesize expert opinions on national defense problems, it has been extended to medical problems (232,246,250,318,336).

A unique feature of the Delphi technique is that persons selected to participate in the process generally have no direct contact with one another. Instead, participants are provided with a summary of the questionnaire responses, usually by mail. Personality or status variables, thus, have little chance to exert influence on a member's opinion, as they might in face-to-face meetings such as the NIH consensus development conferences. By using anonymous feedback, each expert has an equal chance of influencing other participants (41). The technique is also viewed as providing a framework within which to approach the problem in a focused manner. Finally, and perhaps most importantly, the technique provides a limited time frame in which to achieve consensus (41,135). There are a fixed number of iterations, usually three, in the questionnaire feedback process.

Delphi has been used to estimate the probability of an epidemic occurring. Information about morbidity and mortality rates for both the total population and a high-risk population were sought in an investigation reported by Schoenbaum, McNeil, and Kavet (336). The investigators employed a modified version of the Delphi technique using two separate groups of participants. The first group consisted of five experts on influenza epidemiology and virology. Subsequent questionnaires fed back anonymous responses of the participants to the previous questionnaire. The second group consisted of 10 experts in immunization, infectious diseases, and preventive medicine. Their subsequent questionnaires were accompanied by summaries of responses compiled from previous questionnaires. The iterative process was continued until median estimates for each group varied by less than 10 percent from the previous questionnaire's responses. Since results of the Delphi process indicated that the probability of a full-scale epidemic was minimal, subsequent economic analyses revealed that it would not be beneficial to attempt to vaccinate the total population. They concluded that efforts should be directed at immunizing the high-risk population.

The Delphi technique has been criticized as being little better than the "seat-of-the-pants" method currently employed by policymakers, and as being a method which bases "knowledge" on an informal set of opinions rather than on formal decision analysis (332). Others (10) maintain that it is as subject to the same total error found in most predictions. The process is also time and group dependent, since the results are based on the information available to a specific group of experts at a specific point in time. It should be repeated as data change with time. It also appears less well-suited than face-to-face group meetings as a process for resolving minimally controversial issues (318) or for synthesizing the state of the art in a given field (163). Nonetheless, the technique's relevance for gathering predictive information seems clear (77). The Delphi technique may also have use in resolving highly controversial issues likely to be distorted when participants interact personally with one another.

Nominal Group Technique.—In another structured group process, members engage in limited interaction. Typically, all participants may be seated at a common table and asked to write their views on each of a number of issues posed by the leader of the meeting. Each view is recorded on a separate card, and talking is prohibited. The cards are collected, and their contents are listed for all to see without any indication of who is the author of each. The group then discusses these items, often choosing the ones that interest them most. Delbecq, Van de Ven, and Gustafson (82) call this the "nominal group" technique (NGT) because the individuals at the table (at the outset) are a group in name only. The (silent) presence of others while writing the cards creates social facilitation which stimulates participants to do well. Subsequent discussion dwells on the ideas proposed without any likelihood of distraction by attitudes toward those who did the proposing.

Thornell (368) has recently reported a study comparing the Delphi technique with the NGT. Physicians were randomly assigned to one of three Delphi or NGT panels to develop procedures for handling four hypothetical emergency medical services cases. In order to determine the reliability of the decisions, panelists were contacted individually 6 months later and asked to cast an anonymous vote on the procedures originally discussed. The degree of consensus achieved was the same for both techniques. The most striking finding, however, concerned the reliability of decisions over time. There were "very extensive" changes in the NGT vote 6 months later, suggesting that it is "a less than reliable technique for reaching a consensus." In conclusion, although the physicians reported that they liked the NGT much more than Delphi, group norms and pressures were developed with the NGT that produced unstable or false consensual agreement.

Relationship of Assessment Methods to Stages of Innovation

In considering various methodological approaches to medical technology assessment, there are two related issues that must be examined. The first is how to deal with the limited funds available for conducting technology assessments. The second is when or where to intervene in the innovation process. In order to allocate scarce methodological resources, it is necessary to understand some essential properties of technological innovation in medicine.

There are many excellent examples of medical innovations in both the private and public sectors (e.g., 143,252,259). A very recent one—the portable insulin infusion pump—illustrates a number of generic issues in the innovation processes. The case study approach is limited (as was noted above); thus, this example is meant only to be descriptive rather than definitive.

Although the discovery of insulin as a "cure" for diabetes was a major breakthrough, subsequent experience with treatment by subcutaneous injection has revealed that it does not eliminate morbidity or mortality. Currently, diabetes ranks third among major diseases as a cause of death in the United States (309). Moreover, it is associated with a large number of crippling and debilitating conditions. For example, it is the leading cause of blindness. It also leads to myocardial infarcts, strokes, and other serious conditions.

Recently, there has been much discussion of and research on the possibility of using a portable insulin infusion pump to administer and control diabetes (290). Several investigators have demonstrated that such devices control not only blood glucose levels but other metabolites as well (291). Although the exact cause(s) of the various pathologies associated with diabetes are still not understood, the results are nevertheless viewed as significant (133). However, these studies involve few patients—seven and eight, respectively—and should serve only as vivid case studies of the ability of these portable devices to achieve rapid and "strict control" over abnormalities associated with diabetes.

Although the underlying processes of diabetic-related diseases are unknown, it is believed that microvascular injuries (i.e., diabetic microangiopathy) result from the inadequate control obtained by conventional methods, primarily injection. There is now some provocative evidence (133,290) that the strict control obtained with the infusion pump can prevent and perhaps reverse these complications. Thus, there exists a physiological basis or "hypothesis" for the potential efficacy of this device. Such a physiological explanation is often essential for generating interest in a medical technology. When coupled with powerful demonstrations of potential efficacy such as were noted above, a technology possesses the essential ingredients for rapid diffusion.

Two other considerations also figure into the process. The first concerns the safety of the device; the second its availability. As noted above, diabetes is a major threat to human life and well-being. Bunker, Hinckley, and McDermott (40) observed in their review of a number of surgical innovations that under these conditions "efficacy is apt to be considered self-evident." It also appears that safety is seen as nearly negligible in such life-threatening situations. Where there is no alternative treatment and death is the likely outcome, patients and their physicians are motivated to try any promising innovation (382). Under such circumstances, innovations are likely to diffuse and diffuse rapidly. All that is required is sufficient availability or supply of the device. The literature on the infusion pump reveals that there are many manufacturers. It can thus be predicted that this technology is on the theshold of diffusion. Despite the many unanswered questions concerning the long-term effectiveness and acceptability of the pump, despite researchers' claims that it is "an experimental procedure which is still far from being a safe treatment routine," and despite doubts about its effectiveness (350), the ingredients for the rapid diffusion of this technology are all in place.

Type of Technology

Throughout the preceding sections of this appendix, there has been an implicit assumption that the methods described are appropriate for all medical technologies. Is that assumption true? As shown in the preceding discussion, it applies to drugs and surgery, but what about devices, especially those involved in diagnosis?

OTA (266) has described five criteria for assessing diagnostic technologies, one of which is impact on "patient outcome." This criterion has been the emphasis of the methods described in this appendix. Thus, diagnostic devices do not differ in their appropriateness for the methods for technology assessment discussed above. They only differ in the number of other criteria that can be used in their assessment (e.g., accuracy) and in the range of health outcomes they affect.

For example, two of the criteria OTA describes deal with the quality of the information the device provides. This involves established concepts and measures such as specificity and sensitivity of a diagnostic test (see 241). The other criteria deal with the organization and delivery of health services. These are important secondary impacts that should be considered for all technology assessments after the primary determination of efficacy and safety have been made. Banta and McNeil (15) provide an instructive example of these assessment criteria applied to the CAT scanner. They acknowledge that it is difficult to study health outcomes for this type of technology and also difficult to conduct randomized studies of it. As a consequence, secondary impacts involving nonrandomized studies using other criteria may be necessary in the short run. As previously noted in this appendix, such technology assessments require extreme care and cautious interpretation.

In addition to diagnostic, preventive, and therapeutic technologies, OTA (269) considers "organizational" innovations as a major category. Many innovations in health are primarily organizational in their medical function. For example, intensive care units (ICUs) represent a largely organizational change aimed at containing costs by centralizing patient care. Health planners and administrators, in particular, often regard ICUs primarily as an organizational change and not as a well-defined treatment with specified impacts. As Russell (331) notes, it has been "difficult to design a convincing test of intensive care's effectiveness." The confusion between organizational change and health impact has also characterized the movement toward Professional Standards Review Organizations, health systems agencies, and many other major Federal health initiatives. There clearly is a need for planned innovation where the rationale underlying change and its intended impact(s) are specified. The research designs discussed in this appendix are also applicable to these organizational innovations. However, much more attention needs to be given to the implementation processes or operation, and to the integrity of the innovation (340). Thus, it would be important to determine whether emergency medical services have been properly installed before assessing their effectiveness.

A Preliminary Strategy for Assessment

Our discussion of innovation raises the question of its relationship to the various methodologies described in this appendix. A number of researchers, including Williamson (396) and McKinlay,[6] have described models or stages in the innovation process that can be used to relate issues in validity and design to the development of a medical technology. According to a recent NIH conception similar to Williamson's (see table C-2), a medical innovation goes through three stages of development. At the earliest level (i.e., "new"), there is the perception of need such as a cure for a disease or a better way of diagnosing it and a preliminary assessment of the technical feasibility of the idea underlying the innovation (i.e., construct validity). The technology then becomes a reality, usually in an early form (i.e., "emerging") that can be assessed for its efficacy, safety, and social impact (e.g., quality of life). At this point, research design and validity issues (i.e., internal and statistical conclusion validity) as well as cost considerations are important. Once satisfactory evidence is obtained at this level, the innovation develops to an "existing" level where the emphasis is on its acceptability or external validity. Widespread diffusion of the innovation should occur at this point, and the relevant policy issues concern the cost effectiveness of the technology and the continued observation or postmarketing surveillance of the technology for unintended negative side-effects (388). Given the inability to predict the future impact of technologies—especially low-frequency, unanticipated negative side-effects such as toxic shock syndrome—continued surveillance using epidemiological (i.e., case-control) and related methods will be necesssary.

The large number of potential technologies to assess and the pressures to develop and diffuse them quickly ensure that some stages of development will not be scrutinized with the appropriate methods. Many of the above examples (e.g., CABG surgery, gastric freezing) illustrate this point. In fact, it is the overdiffusion of young technologies and their associated costs that have led to the need for strategies to deal with the problems of technology assessment. The model described above, coupled the methods presented, provides the basis for such a strategy. There remains a need to order the technologies according to their priority for systematic, thorough assessment.

According to the model, technologies in the first stage of development do not need to be assessed. Since many, if not most, medical innovations will not progress beyond this point, the burden of assessment will be considerably reduced. Technologies maturing beyond this level can be ordered by the potential benefits and harm they pose. This ordering could be determined simply by calculating the product of the benefit or risk the technology poses to either decreased or increased mortality multiplied by the amount of use envisioned for the technology. For example, CABG surgery may pose a 4 percent risk of death for the 100,000 patients operated on last year. This would result in 4,000 deaths. Another decision rule could involve cost. Obviously, medical technologies could be ordered by both of these rules. The choice among the various possible ordering procedures is one that falls in the policy domain and is beyond the scope of this discussion.

Conclusions

A brief examination of innovation in medical technology reveals that it is a dynamic, temporal process that requires considerable flexibility in the methodology used. Different approaches are relevant at different stages of technological development. Moreover, policy-relevant evidence may not be available when needed, either because of the pressures for diffusion or the low priority for assessment initially assigned to the innovation. The Dalkon Shield, an intrauterine device, is a recent, unfortunate example of premature diffusion. Furthermore, no matter how thorough the assessment of a medical technology, there is always the possibility that unanticipated negative side-effects will be discovered at a later date when use of the technology is more widespread (e.g., X-ray treatment for facial acne). At such times, a decision to reexamine the technology will have to be considered along with the choice of an appropriate methodology for accomplishing this. Such postdiffusion technology assessments are much more difficult to accomplish. The recently initiated RCT to assess chemotherapy as a treatment for osteogenic carcinoma is an example of this process of surveillance and reassessment.

Given the scarcity of resources, it is unlikely that there will be much increase in the number of large-scale RCTs. Most of these require Federal support, and considerable funds are already allocated for such technology assessments. For example, Levy and Sondik (222) report that in 1976 about one-eighth of the total budget for NIH's National Heart, Lung, and Blood Institute—over $50 million—was devoted to major clinical trials. However, for many innovations, small-scale, local RCTs are probably feasible. Unfortunately, these are not often conducted. One can only speculate as to the reasons for this. Physicians often lack the methodological training to conduct such studies or the conviction that single-site studies are useful. The implication for medical technology assessment is that there will be an increased reliance on studies using other methods of evaluation unless some new policy initiative (see 307) is taken. As noted in this appendix, these other evaluative designs are most vulnerable to challenge and are often seriously flawed. Where such quasi-experimental approaches are employed, replication and "triangulation" (71)—the use of multiple lines of evidence to eliminate or reduce the salient threats to validity—should be encouraged.

---

[6] J. B. McKinlay, "From 'Promising Report' to 'Standard Procedure': Seven Stages in the Career of a Medical Innovation," *Milbank Mem. Fund Q.* 59:374, 1981.

When should one conduct a large-scale RCT? Levy and Sondik (222) describe a complex multiphase, multigroup decision process based on four broad decision criteria: knowledge, methodology, resources, and ethics. Methodological considerations, involving power, significance level, effect size, and the like, are used to estimate the number of subjects and the length of the study. These factors determine the cost of the study and hence its feasibility. In sum, Levy and Sondik outline a complex group decision process that provides a type of cost-benefit analysis for conducting an RCT. The emerging methodology of decision analysis (386) would be useful in selecting medical innovations for such high-quality technology assessments.

In conclusion, the dynamic nature of medical innovation requires constant monitoring. This can be accomplished either through postmarketing surveillance (as noted above) or by careful, systematic reviews of the accumulating literature dealing with the innovation. Thus, medical technology assessment must not be viewed as a one-time event. As the model described in table C-2 indicates, evaluative studies for technology assessments should be considered at all stages of development, particularly during the second stage.

Appendix D.—Medical Technologies and Innovation

Introduction

In some respects, the innovation process for medical technologies parallels that for other technologies. Although there are many variations, the basic process is as follows. An innovation is conceptualized by recognizing both technical feasibility and potential demand. If a decision is made to pursue the innovative idea, problem-solving activity follows, drawing from available information and further research and development (R&D) activities. If a solution to the problem is found, it may be the one originally sought, or a solution to a modification of the original problem. The final stage before widespread utilization of an innovation is its introduction into the market.

It is at this point that the innovation process for medical care technologies differs from that for most other technologies. Drugs must meet premarket approval requirements for efficacy and safety. Medical devices, depending on their classification, must either meet general controls, adhere to performance standards, or meet premarket approval requirements for efficacy and safety. New medical and surgical procedures, though not subject to the same regulatory requirements as drugs and devices, are increasingly subject to more systematic applications of clinical testing to evaluate their efficacy and safety; and decisions to pay for their use are also increasingly being subjected to more systematic analyses by private and public health insurers.

Definitions of Innovation

The basic criterion for an innovation is "newness," or "differing in significant ways" from previous products or programs (213). In its most limited definition, an innovation is an invention that is regarded as novel, independent of its adoption or nonadoption (405). But in other definitions, inventions are not considered innovations unless the adopting system perceives them as such—i.e., innovation involves the process of conceptualizing a new idea, finding the solution to the problem, and *using* a new item of economic or social value (256).

These different concepts of innovation impinge on the question of whether regulatory and reimbursement policies inhibit the innovation process. One might find, for example, that regulation reduces the number of new patents. Using the invention concept of innovation, one might then conclude that innovation has been hindered. Patents, however, offer little insight into the value of inventions. Even innovations that have achieved widespread use are not necessarily beneficial (252). From that standpoint, inventions that do not show social utility as well as economic worth are not innovations. Thus, it could be argued that inventions that do not meet regulatory or reimbursement criteria (representing collective judgments on social utility as well as economic worth) are not innovations.

Research on Innovation

There are four principal approaches of research for understanding the innovation process: 1) statistical studies, 2) contextual comparisons, 3) critical incident studies, and 4) case studies.

National level statistical studies concerning innovation might include the contribution to gross national product, rates of diffusion, effects of legislation, etc. Such studies suffer principally from a lack of differentiation. Innovations vary enormously in terms of complexity, radicalness, compatibility, etc., and organizations and industries vary in size, technology, history, culture, etc. However, most statistical studies generalize on the basis of an assumed homogeneity.

Contextual comparisons are made by selecting and comparing organizations that are similar along several dimensions (e.g., size, technology, product range) but differ in terms of success at innovating (or some similar dimension). From contextual comparisons, it may be possible to extract a list of factors common to the successful innovators but not to the others. Given large samples, the regularity with which some factors appear confirms their importance in the innovation process. Contextual comparisons cannot account for all the local sources of variation, however, and remain at a general and somewhat superficial level.

Critical incident studies deal with individual recollections about important stages in the development of various innovations. The problem here is one of subjective emphasis and bias, rich in detail but not necessarily giving the whole story. Critical incident studies also tend to deal with major innovations only and to provide little information about incremental changes, the total contributions of which may equal or exceed the contribution of single radical innovations.

Case studies are a frequently used approach in medical technology assessments. An attempt is made to get close to the process for a long period of time from an involved but neutral viewpoint. Case studies represent an attempt to understand the dynamics of a process which is naturally changing in character and content all of the time, something very few of the other types of studies consider. But case studies lack a developed methodology, and, although they may be able to account for the behavior of one specific organization, there is no basis for generalizing beyond that to others.

The limitations of these four basic approaches are: 1) they either provide general data which are limited in applicability to specific circumstances; or 2) they give a highly specific account of one organization or invention, identifying most of the factors which influence the innovation process but with no direct general applicability (27).

Since such studies attempt to place a rational, predictive framework on creativity, it is not surprising that they provide an enormous amount of descriptive detail but do little in the way of establishing cause-and-effect relationships. Innovation seems to be the result of many interrelated factors and not of any particular factor. Nevertheless, it is useful to review the available research findings to help gage what effects regulatory programs and changing governmental reimbursement policies can or might have on the innovation process.

The next section of this appendix summarizes what is known about the factors that affect the innovation process for drugs, devices, and medical and surgical procedures. The second section discusses regulatory mechanisms and medical care reimbursement policies and draws inferences concerning their possible effects on the innovation process.

Factors That Affect the Innovation Process

Characteristics of Successful Innovations

When successful innovations are examined for their key characteristics, certain recurring factors are commonly found in all industrial areas. Their relative importance varies from industry to industry and even between specific innovations in one industrial area, but together these factors provide a composite picture of the conditions under which the innovation process thrives.

Personnel of five types contribute to successful innovations (313). "Idea-exploiters" (as opposed to "idea-havers") not only think up new ideas but also do something about them. "Entrepreneurs" (or "product champions") advocate and push for change and innovation. "Program managers" (or "business innovators") handle the supportive functions of planning, scheduling, business, and finance related to the developmental activities of their technical colleagues. "Gatekeepers" (or "special communicators") are the links who bring information from outside sources, joining technical, market, and manufacturing sources of information to the potential users of the information. Finally, "sponsors" or ("coaches") are senior people not carrying out the research or advocating the innovation, but providing junior people with the resources necessary to move technological advances forward in the organization.

Motivating forces for the initiation of innovative activity are roughly divided into "technology-push" and "market-pull" theories. The former reflect the belief that pushing technology through basic research will eventually result in significant technological development. The latter reflect the belief that the market, through recognition of a need and creation of a demand for new products, is the dominant factor in producing successful innovations.

The general industrial literature supports the theory that market-pull is the primary influence. From 60 to 80 percent of important innovations across the industrial spectrum have been related to market demands (375). In a study from West Germany, 70 percent of successful innovations originated from market-pull and 80 percent of failures began with technology-push (156). However, it is apparent from this literature that it is not an either/or situation between technology-push and market-pull.

Comroe and Dripps (64) argue that in the area of biomedical technology, technology-push is a more important factor. In studying the 10 most important clinical advances against cardiovascular and pulmonary diseases from 1945 to 1975, these investigators reviewed 529 publications considered to be the key research articles leading to these advances. They concluded that 41 percent of the key articles "reported work that, at the time it was done, had no relation whatever to the disease that it later helped to prevent, diagnose, treat, or alleviate."

As for *sources of effective technical solutions:* "In most industries, no single firm commands a majority of the resources available for research, nor can any one firm respond to more than a portion of the needs or problems requiring original solution. It is not surprising, therefore, to find that most of the ideas successfully developed and implemented by any firm came from outside that firm" (375). Moreover, the predominant route of information is personal experience and contacts, not the scientific literature.

An effective technical solution may be an original innovation or one adopted or adapted for a particular problem. About 20 to 30 percent of significant innovations are adopted or adapted ones (219,256) and, as might be expected, a new technology has a greater propensity to be adopted or adapted for a new use when it has passed through the initial and developmental stages into the late maturity stage (376).

Finally, the user or the manufacturer may be the source of the solution, and studies have shown that in many industries (e.g., computers, specialized machinery, scientific equipment), a user came up with the solution, which was then adopted and turned into a product by the manufacturer. Roberts (313) believes that in the medical devices industry, the manufacturer's role is primarily one of adoption and broad-based distribution.

Channels for exploitation is the stage that precedes widespread diffusion. In medicine, the typical mechanism is the clinical trial, and the evidence concerning whether clinical trials function effectively to transfer research results into clinical practice is conflicting (223, 404). In most industrial fields, the nonprofit sector contributes infrequently to innovation. In biomedical innovation, however, universities, medical schools, and hospitals are crucial. Yet few linkages exist in the biomedical area between academia and industry to put innovations into widespread use through commercial marketing. Some linkages may result from a recent change in the patent laws (Public Law 96-517) to enhance commercial exploitation of inventions developed with Federal assistance. And in genetic engineering, universities, medical schools, and hospitals are now forming business relationships with industry, receiving substantial amounts of research funds in exchange for exclusive licenses to market the anticipated innovations.

In the *lifecycle of a technology*, major technological changes occur in the early stages, but incremental technological changes usually dominate the later stages (376). In other words, product innovation dominates in the early stage, with little change in manufacturing process; but as the technology progresses, there is a rapid decline in product emphasis and dramatic increase in process orientation. Finally, small companies contribute most to innovation in the early stages of a technological field, but large companies dominate by the time the field matures. Roberts (313) has observed this pattern in genetic engineering, a new field where small companies are the dominant contributors.

Innovation Process for Drugs and Devices

Although there is little information specifically related to economic factors that affect innovation in the drug and medical devices industries, the following general observations are probably applicable. Innovation is one way to compete in the market and is part of some companies' overall strategy. Basic questions are the extent to which a company pursues long-term v. short-term strategies, and the extent to which competitive success over the long run depends on a company's commitment to technological superiority.

Mansfield, et al. (235), define three probabilities for assessing the importance of different factors at different stages of the innovation process: 1) the probability of successfully completing the technical problem-solving stage; 2) the probability of successfully completing the commercialization stage, given that the technical problem-solving stage has been completed; 3) the probability of economic success, given commercialization. (Economic success means that the project will yield a rate of return which is equal to or in excess of that available from alternative investments.) The product of these three probabilities is the probability that a project which is initiated will be an economic success.

The aforementioned probabilities are affected by both external and internal factors. Externally, a high rate of inflation means high interest rates. These, in turn, make it more expensive to raise capital for long-term investments, and produce large fluctuations in prices, thereby garbling the relative price signals which producers use to determine the kinds of production processes that would minimize future costs. These uncertainties might turn corporate strategy toward a short-term focus.

Corporate strategists also have to choose between long-term technological breakthroughs and short-term, quick-payback product and process improvements. A major innovation of great technical novelty may not have a well-defined market potential. In contrast, a modest product improvement may have a highly predictable market. The major innovation may have a much greater profit potential, but the risk of failure is also much greater. As mentioned above, one study found that 70 percent of successful innovations originated from market-pull and 80 percent of failures began with technology-push (156). Projects that originate with R&D personnel are more likely to be technically challenging than projects that originate with marketing or other company personnel. They also have a lower probability of successful commercialization, because R&D personnel are likely to have less understanding of market potential. But once past the commercialization stage, projects originating with R&D personnel have a higher probability of economic success (235), presumably because they are the most likely to have the combination of technical and economic factors necessary for ultimate success in the market. Whether the greater probability of economic success can offset the lower probabilities for technical completion and commercialization is a matter of judgment. Many projects are not initiated or carried through to completion, because they are judged either to have insufficient market potential or to have risks that are too great.

There is evidence that corporate strategy in the United States has turned increasingly to a short-term focus, and opinions have been expressed that this is the primary source of our current problems concerning innovation, declining productivity growth, and balance of trade with other countries. Recent investments have been skewed toward equipment and relatively short-term projects and away from structures and relatively long-term investments (231), and an increasing portion of industrial R&D is directed toward relatively short-term developmental work and less toward long-term fundamental research (257). In this vein, some critics have accused corporate managers of relying too much on near-term market considerations in selecting R&D projects. These critics caustically recall that "the initial market estimate for computers in 1945 projected total worldwide sales of only 10 units" (183).

These critics also contend that current management practices in the United States lead to focusing on short-term, low-risk projects. For example, the decentralization of organizational structures requires a greater dependence on short-term financial measurements, such as return on investments, for evaluating the performance of individual managers and groups. In addition, compensation plans for company executives reward shortsighted behavior. In a survey of 174 companies, 79 percent rewarded executives for short-term performance, and only 42 percent offered "long-term" incentives, which were defined as anything over 1 year. In another survey, year-end bonuses were larger than long-term incentive awards; while bonuses amounted to about 50 percent of salary, the median long-term award was only 34 percent of salary (298).

Finally, there is an increasing proportion of corporate presidents with legal or accounting as opposed to technical backgrounds. Critics maintain that this trend reflects a shallow concept of the professional manager as "an individual having no special expertise in any particular industry or technology who nevertheless can step into an unfamiliar company and run it successfully through strict application of financial controls, portfolio concepts, and a market-driven strategy" (183). Such critics contend that although technological issues must be an integral part of broader strategic issues, they cannot be handled by the same methods applied to finance and marketing.

Regardless of the industrial sector, most small manufacturers do not engage in formal R&D. For firms undertaking R&D, innovational effort tends to increase more than proportionately with firm size up to some point that varies by industrial sector. Innovations produced mainly by large firms are typically those in capital-intensive industries. The exceptions are in aerospace, shipbuilding, and pharmaceuticals, where capital intensity is low but development costs for new products are very high. These findings are illustrated in the United Kingdom. Between 1945 and 1970, small manufacturers produced none of the 44 innovations produced by all U.K. pharmaceutical firms, but small manufacturers' share of net pharmaceutical output in 1963 was 12 percent (151,327).

The U.S. drug industry is also characterized by high and rising development costs for new products and a strong shift toward greater concentration of new products in the very largest of the approximately 600 pharmaceutical firms. Since the late 1950's, the number of firms producing a new chemical entity has declined, and the development of new chemical entities has been increasingly concentrated in the top four and eight largest firms (see table D-1). In other words, innovative outputs have been concentrated in the 20 largest of the 600 drug firms, and most of this concentration is among the top four to eight innovators.

While the four largest firms' share of innovative output remained stable from the late 1950's through the early 1960's, then accelerated sharply, their share of total prescription drug sales remained fairly constant (see table D-2). Taken together, these findings indicate that the increasing concentration of new chemical entity output in fewer firms has accrued to large firms, mostly at the expense of the smaller firms, in the top 20 innovators. But the four largest firms, despite a near doubling of their share of innovative output, have had essentially the same share of total prescription drug sales during this period. Most of the large drug firms are dependent on a few drugs for much of their income. For example, the three leading products of the companies listed in table D-3 accounted for 22 to 84 percent of their total U.S. pharmaceutical sales in 1979.

These observations probably reflect the following scenario: The vast majority of the 600 U.S. drug firms

Table D-1.—Concentration of Innovational Output in the U.S. Drug Industry, 1957-71

Period	Total number of new chemical entities (NCEs)	Number of firms having an NCE	Innovational output of concentration ratios		
			4-firm	8-firm	20-firm
1957-61	233	51	0.462	0.712	0.931
1962-66	93	34	0.546	0.789	0.976
1967-71	76	23	0.610	0.815	0.978

SOURCE: H. G. Grabowski, et al., "The Effects of Regulatory Policy on the Incentives To Innovate: An International Comparative Analysis," 1976.

Table D-2.—Percentage of Innovational Output and Total Sales Accounted for by the Four Largest U.S. Drug Firms, 1957-71

Period	Four largest firms' share of innovational output	Four largest firms' share of total drug sales
1957-61	24.0	26.5
1962-66	25.0	24.0
1967-71	48.7	25.1

SOURCE: H. G. Grabowski, et al., "The Effects of Regulatory Policy on the Incentives To Innovate: An International Comparative Analysis," 1976.

Table D-3.—Percentage Total U.S. Pharmaceutical Sales Accounted for by Three Leading Products, Selected Corporations, Selected Years[a]

	1970	1975	1979
Abbot.	36	33	28
American Home Products:			
Ayerst	64	74	84
Wyeth	37	44	43
Bristol-Meyers:			
Bristol	69	46	28
Mead-Johnson	40	38	37
Burroughs-Wellcome.	NA	56	51
Ciba	47	NA	55
Lederle	48	31	32
Lilly	46	60	43
Merck	35	44	44
Pfizer	52	65	65
Robins	43	45	46
Roche	80	80	70
Searle	45	49	44
Shering	42	48	40
SmithKline	44	42	66
Squibb	28	31	23
Upjohn	47	50	56
Warner-Lambert:			
Warner	53	NA	NA
Parke-Davis	25	27	22

NA = not available

[a] U.S. sales

SOURCE: Charles River Associates, Inc., "The Effects of Patent-Term Restoration on the Pharmaceutical Industry," a report to OTA, May 4, 1981

are small manufacturers producing primarily generic drugs for limited markets, but also other patented drugs. After patents expire, generics erode some of the market captured by large innovator drug firms and these firms regain their share of total sales through the introduction of new drugs.

The *U.S. medical devices industry* has experienced substantial growth since World War II. Industry sales in 1977 were $8.1 billion—five times the amount in 1958 (corrected for inflation). Growth has been predominantly in the number of firms rather than in their size. The U.S. medical devices industry is composed of several thousand firms—many specialized small firms which together have a small share of the market and a few large firms with high market shares. There are high entry and exit rates in the industry, mostly among small firms (8). Profitability is higher than average in the economy.

Dominance by large companies suggests the presence of economies of scale, while the persistence of many small companies suggests that economies of scale do not apply to specialized areas. Possibly, however, the large firms really represent the industry; i.e., rather than representing the differentiation of the industry into small and large functions, the large number of small firms may represent a high-birth, high-mortality, and high-turnover sector of the industry (122). Arthur Young & Co.'s survey of the industry, for example, did not differentiate between bankruptcy and acquisition in its observation of the high-turnover rates for small firms. However, D'Arbeloff (79) comments that high-turnover rates may reflect a high-risk, high-profit atmosphere for small firms.

In general, small firms fill a special niche in the medical devices market, and their growth into larger firms is hindered by conditions such as advertising requirements, links with distribution channels, and the need for new capital expenditures (355). Thus, the industrial pattern is that of limited internal growth, with acquisition or establishment of smaller companies being the primary method of expansion. Small plants are opened to manufacture new products following invention and development, while large plants are opened by large companies to take advantage of lower operating costs. These large companies tend to be extremely diversified as a whole, yet there is little product diversification within their medical devices plants (8).

Recently, the distribution of medical devices has shifted from small regional and local suppliers to major national dealers. National dealers are often subsidiaries of large manufacturers or are acquirers of small manufacturing firms. The advantage of larger firms is that they are better positioned to provide special buyer education through their larger, better trained staff (355). The inability of potential manufacturers to gain access to these networks is an additional barrier to growth of the small firms entering the medical devices field and probably accentuates their acquisition by larger manufacturers.

The U.S. medical devices industry is somewhat insulated from price competition by the high level of third-party reimbursement, and price competition is not as significant a force in mitigating price increases as it is in other industries. Nevertheless, there is a high degree of product differentiation, and the industry appears to be competitive at various levels even though the market for the most part is price insensitive (8). In other words, a policy of product differentiation and sales promotion may increase a firm's net revenues above the competitive level (288). Profitability measures confirm this viewpoint, indicating a slightly higher profitability in the devices industry than exists throughout the economy. This may explain the observed trend of expansion through acquisition (product differentiation), coupled with major national dealers (sales promotion) either being subsidiaries of large manufacturers or being acquirers of small firms. Product differentiation, distribution, and perhaps the level of new capital investments also appear to act as inhibitors on the growth of small firms (8) and contribute to such firms' failure or acquisition by larger firms.

Innovation Process for Medical and Surgical Procedures

The invention, development, and diffusion of medical and surgical procedures may generally be described by the model of the innovation process developed for products and their manufacturing processes. New procedures usually involve some drug and/or device, and innovations in medical and surgical procedures can be viewed as user-generated innovations, where a previous innovation is adopted or adapted (modified) for another purpose. Regardless of how medical and surgical procedures fit into the model of the innovation process, however, a focus on procedures separate from the drugs and devices that are used in them is necessary, because physicians, as users, are both generators (technology-push) and purchasers (demand-pull) of innovations. Thus, it is crucial to get at least a notion of how they perform these dual roles. But there are no standard determinants of when or how procedures become medically acceptable (197) and few criteria for when they become obsolete.

There are three separate literature sources for analyzing the dissemination of information in medicine. The first comprises sociological research on the diffusion of innovations in social systems (208,317); the second is literature concerned with the effects of communication variables on attitudes and behavior (239); and the third is the scattered, nontheoretical literature in medicine, consisting of descriptive studies of the dissemination and adoption of different medical innovations (62,145,233,331).

The medical literature on the dissemination and adoption of innovations is weighted toward studies of single medical technologies which are diagnostic or therapeutic in purpose. There is a large literature on how physicians learn about and adopt new drugs and a growing literature on specific devices or techniques, but little is known about communication about or the adoption of complex medical procedures which may not involve drugs or hardware (e.g., psychotherapy).

In practice, however, the crucial distinction is between communication which informs physicians about novel technologies and that which influences physicians to act (405). Even though the most important source of new knowledge about improvements in medical technologies is the professional literature, physicians cite professional colleagues more often as sources they turn to when contemplating actual implementation of new procedures (145,233,234).

The importance of informal communication both in the process of scientific discovery and in the diffusion of technological innovations seems to be a feature not only in medicine but in all fields of technological discovery and diffusion (213). Moreover, it may be that there is a prestige hierarchy in which those at the top are "trend setters" (49). If this is so, widespread adoption of an innovation could be enhanced by convincing influential organizations to adopt it first, then letting prestige-seeking organizations imitate them (213).

Physicians of greater prestige do tend to hear about innovations sooner than others (62), and they are also mentioned by their fellow professionals as influential sources of information on the medical practice of others. However, the adoption process when the adopting unit is an organization (e.g., hospital) is substantially different from the process when the adopting unit is an individual (e.g., physician in solo practice) (178, 405), and these processes differ by the level of complexity of the organization. Outside forces such as third-party reimbursement or regulatory practices may also affect how quickly the individuals in the medical community learn about or adopt a technology.

The following general scenario may help make these theoretical and empirical findings more concrete. Medical and surgical procedures usually begin as user-generated (e.g., physician) innovations. In medicine, an innovative procedure may be in the form of adopting an existing drug for a new purpose or changing the mixture of drugs and their dosages to adapt them to a different medical problem. In surgery, it may be in the form of a modification of an existing technique (usually in accompaniment with modifications of the devices being used) for application to a new use. In treatment areas that do not depend on drugs or devices (e.g., psychotherapy) or in which drugs and devices are used but are not crucial to the innovation (e.g., primary care), it may be an innovative interpretation of the existing knowledge (e.g., the multiple schools of psychotherapy which have sprung up, the "family physician").

Increasingly, innovations in procedures arise in academic or academic-associated centers, where physical and professional resources are readily available; a research, innovation-seeking atmosphere is encouraged; and contacts with others in the field extend not only nationally, but also globally. Innovators in such settings know how to present the innovations in a manner that will be technically acceptable, and they also have the prestige which gives them access to professional meetings and journals to publicize their results. Their presentations and publications not only diffuse the innovation to a wider audience, but more importantly, begin to legitimize it. Depending on the claimed innovation's nature, usually defined in terms of how the innovation will revolutionize or at least substantially affect the related area of medical or surgical practice, other academic centers will begin to pursue it, too.

At this point, several Government agencies may enter the picture. The National Institutes of Health (NIH) may provide support for the innovator and researchers in other health centers in the form of randomized clinical trials (RCTs), most likely conducted in some of the clinical research centers funded by NIH. A new use for a drug, invention of a new device, or modification of an existing device requires the Food and Drug Administration's (FDA's) approval. Increasingly, investigational new drug or device uses approved by FDA for limited testing are given to the same centers which NIH supports as clinical research centers (or at least to the health institutions in which these designated centers are located). Sooner or later, the Health Care Financing Administration (HCFA) may receive a request for reimbursement of the new procedure and will give great weight to the NIH clinical trials for evidence of safety and efficacy. Meanwhile, FDA must make a determination of safety and efficacy for market clearance of the drug or device under review. FDA will often have to make its decision long before NIH reaches a decision and terminates funding for the clinical trials. The reason is that FDA must act in a timely manner and reach its conclusion on minimal evidence, while NIH has no similar regulatory responsibilities and is more interested in the cumulative evidence. FDA's decision, moreover, especially in the case of devices, may rest on the narrow question of the efficacy and safety of the device in a particular setting, not of the entire procedure in general use. But release of the device to the general market, once premarket approval is given, also tends to speed up the diffusion of the procedure which NIH may be studying. This result, in turn, places more pressure on HCFA to reimburse for the procedure.

Most of these points are illustrated in the brief case studies in appendix E on: 1) gastric freezing for the treatment of ulcer, 2) hemodialysis for the treatment of schizophrenia, 3) percutaneous transluminal coronary angioplasty, 4) maternal serum alpha-fetoprotein, and 5) hemodialysis and kidney transplantation.

Funding of the basic research which advances medical care comes primarily from NIH, with smaller but important amounts from private foundations (223). The central role which basic research plays in the process of medical innovation (64) is the justification for the substantial public and private moneys invested.

In the development and diffusion phases of medical innovation, initial findings are translated into clinical procedures. These phases are central to the innovation process, but there is relatively little formal funding. The National Center for Health Services Research (NCHSR) was originally called the National Center for Health Services Research and Development, but its enabling law, when finally passed in 1974, specifically forbade the center to fund development. Although about half of NCHSR's grant awards have been for projects classified as demonstrations, little has been devoted to new medical and surgical procedures.

The primary focus of NIH is research, and there appears to be no systematic or comprehensive policy of NIH support for development. Figures to document the size of NIH's investment in development are not available. Although NIH grants and contracts have been given to support development in a number of areas (e.g., the artificial heart program, cancer screening, cancer chemotherapy, and, in recent years, hemodialysis), the amount invested in development probably constitutes a relatively small portion of the current $3.8 billion NIH budget.

For developmental costs of procedures used in the prevention or treatment of individual diseases, private foundations have provided important support. A notable example is the generous funding by the Hartford Foundation of Dr. Belding Scribner's hemodialysis program in Seattle in the early 1960's. Other examples include grants by the American Cancer Society for cancer screening and treatment programs and by the Jules Stein Foundation for the development of radial keratoplasty (a type of surgery on the eye).

Although there are no explicit data on which to base estimates, the developmental costs of medical innovation are without doubt very large. By and large, the costs of the developmental phase of early clinical application have been paid by patients, usually through standard medical insurance policies.

Even for procedures that have been clearly designated as experimental, reimbursement has often been provided. Thus, for example, when total hip replacement was first introduced into this country in 1971, it came under the aegis of FDA because of the use of the acrylic, methylmethacrylate, in the operating room construction of the new joint. Despite the artificial hip's clear designation as an experimental device by FDA, the total hip procedure was reimbursed from the outset as an acceptable surgical procedure.

Heart surgery has similarly been reimbursed from the outset through standard medical insurance policies. The single exception was the introduction of heart transplant surgery at Stanford in 1969. Other institutions performing heart transplants have simply charged standard fees. Coronary artery bypass graft (CABG), when introduced in 1969, was considered by its innovators to be standard therapy, despite repeated calls for randomized clinical trials (RCTs) of the new operation (52,358). Clinical charges for CABG were paid via standard policies from the outset and continued to be paid even when, several years later, CABG was finally subjected to RCTs.

Benson Roe, a cardiac surgeon at the University of California School of Medicine in San Francisco, has recently described the historical justification for the "extraordinary" fees in cardiac surgery (316):

> Historically, of course, there was justification for extraordinary fees in cardiac surgery. The developmental years of this field were indeed difficult, demanding innovative talent and an enormous amount of time—requirements that many were unable to fulfill. The early cardiac surgeon participated in the diagnostic studies and preoperative preparation, planned and directed the technical details of the cardiopulmonary bypass, conducted the entire longer operation, and personally supervised every detail of postoperative care, often spending late nights at the bedside.

As heart surgery has become, if not routine, at least a great deal more safe and considerably simple, Roe suggests (316):

> . . . one might expect the surgeon's fee to have dropped considerably, but it has not. On the contrary, fees for cardiac surgery have escalated at a rate that far exceeds the inflationary factor.

Much the same pattern Roe has observed in the case of cardiac surgery has been followed for other technologically complex surgical procedures, including intraocular lens implantation and microdissection in brain surgery, as well as orthopedic joint replacements. Not only are the enormous costs of medical and surgical development absorbed by medical insurers (222)—and eventually by the public—but the charges for new procedures, once standard, remain high.

Public Accountability

Regulatory actions and more informed reimbursement decisions are intended to help ensure that new and emerging technologies are efficacious, have acceptable risks, and are appropriately used (e.g., are cost effective). Private industry determines which drugs and devices it will develop primarily on the basis of market-based criteria. To address perceived deficiencies of the market approach, governmental actions infuse additional criteria based on social and political concerns.

These governmental actions have generally been regulatory in nature, concentrating on the costs to our health, safety, and environment—costs which, because they are diffuse, can best be addressed through collective, governmental actions. Government's role as a purchaser of technologies, of great significance in health care because of Government's role as insurer, has also led to a need to make more informed judgments about the kinds of technologies used in health care. These judgments are needed not only to minimize reimbursing for the use of ineffective technologies, but also to help decide which among the array of technologies are the most appropriate. The regulatory process unquestionably slows diffusion of technologies into the marketplace, and some technologies are filtered out. Slowing the diffusion of new technologies may allow for more informed and timely decisions before widespread use.

Constraining the diffusion of new drugs or devices before they are adequately assessed also affects the conditions under which new technologies are fostered. Meeting regulatory requirements for evidence of efficacy and safety increases industry's costs, for example, by delaying industry's return on capital invested in R&D activities. Factors such as these play a significant role in industry's assessment of whether a new technology could be profitably marketed or in deciding which of several promising technologies to develop further. But the full extent of a new technology's capabilities is usually not known until it is put into use, and use can lead to improvements and, in some cases, further innovations.

The question of the effect on innovation from regulatory and reimbursement policies is not simply one of whether innovation is inhibited, but also whether the alterations in the innovation process are unintended and undesirable. Government support of R&D has long sought to alter the innovation process, most notably to accelerate the pace of innovation and to push it in certain directions. NIH is a prime example of both undirected and directed support for the development of new medical technologies, combining basic research within separate institutes targeted at specific diseases.

As the recent experience of air quality control programs demonstrates, market-modifying factors such as regulation can also alter the direction that innovations take. The kinds of regulations put into effect can force innovation along certain pathways, some of which allow for more maneuvering (e.g., in contrast to specifying the kinds of pollution control devices to be installed to achieve air quality standards, the "bubble" concept of regulating air pollution sources, where a maximum air pollution level is set, leaves it to pollution sources to stay within those limits by whatever techniques they can muster). Restraints on the marketing end of the innovation pathway confront innovators with new conditions, and the hallmark of innovation is to generate new answers when conditions change. Although there will be industrial losers and winners under these regulatory rules, that does not necessarily mean that innovation has been hindered. It may instead have its direction altered, much as Government attempts to alter innovation at the R&D end for similar social purposes.

There is general agreement that competition among medical care providers is typically not based on price (331); under current reimbursement policies, there are incentives to adopt all available diagnostic tools and to pursue any therapy anticipated to have any value. This is particularly true for hospitals. Third-party coverage currently accounts for about 90 percent of expenditures for hospital care. As the price of technology has little effect on providers and patients under existing health insurance arrangements, a greater adoption of technology can be expected to occur under these arrangements than would occur under more price-competitive reimbursement arrangements.

At a simple level of comparison, recent changes in current regulatory and third-party reimbursement policies can be thought of as approaching some middle ground from opposite ends of the spectrum. Regulation purposefully slows down the innovation process, particularly at the early diffusion stage, and modifications are now being sought (e.g., in premarket approval requirements for drugs) to ensure that this slowing of the innovation process is no more than necessary to achieve the regulatory program's objectives. Current reimbursement policies, on the other hand, are seen as boosting the diffusion of new medical technologies and modifying existing technologies beyond what would take place under more price competitive systems, and reforms are being aimed at constraining the adoption process.

Because the purpose of regulation is to infuse social criteria into judgments of a new technology's worth, conclusions based on the economic impact of regulatory requirements must be reached with caution. Regulation is expected to change the innovation process. The issues are whether the specific changes were intended and whether the benefits of regulations are worth the price paid in resulting alterations of the innovation process.

Present reimbursement policies tend to reward the use of technological innovations and discourage less technologically oriented patient care activities (2). Thus, there is a need to infuse more price sensitivity into the reimbursement system. Taken together with the regulatory approach, changes to infuse price sensitivity would theoretically: 1) allow market entry of innovations which have met social criteria of worthiness, and 2) make it possible for those new technologies which have passed the regulatory test to then compete with one another on a price basis. Curtailing excessive demand by a more price-sensitive approach, however, means changing the conditions of the current medical technology innovation process. Again, the question here is whether such major changes in the demand for new medical technologies will affect the innovation process in unintended and undesirable ways.

Regulation

The purpose of regulation is to guide the course of technical change in such a way that, over time, new technologies are responsive not only to the cost and performance characteristics valued by the marketplace, but also to the social values that motivate regulation. Regulatory requirements become added conditions for successful completion of the innovation process.

There are three possible approaches to regulation:

1. *precluding* technologies deemed socially undesirable by either banning or selectively restricting their use;
2. *deflecting* technologies by forcing their development or diffusion (e.g., through uniform requirements) into technologies with performance characteristics deemed socially desirable; and
3. *using market-like mechanisms* (e.g., pollution fees or marketable pollution rights) to encourage producers to economize on the use of common resources such as air and water (244,283).

Of these regulatory constraints, preclusion and deflection are the methods currently used to regulate biomedical technologies. Prescribing how a product is to be made is preclusive, while specifying the qualities the product must have is deflective. The difference is between standardizing the product (preclusion) and standardizing its performance (deflection).

In addition to these purposeful constraints, regulation requires compliance outlays and introduces a number of other factors which can indirectly constrain the innovation process. Compliance outlays include such direct costs as efficacy and safety testing, legal fees, and employee time spent on regulatory matters.

Greater R&D costs are usually associated with the more technically demanding regulations, such as those applied to drugs. These are resources that could be spent in other areas (e.g., development and marketing) or could enhance profit margins.

Uncertainty discourages risk-taking and prolongs decisionmaking, and regulation can introduce uncertainty over how to comply with the regulatory requirements, which may constitute a "moving target." For example, additives are not allowed in foods if they are found to cause cancer. But technical advances in detecting smaller and smaller concentrations of one substance in another (in some cases, at the level of 1 in 1 billion), coupled with the regulatory interpretation that any amount detected is illegal, mean that complying with the law depends on the latest advances in detection methods, even if the best method of keeping the banned substance out of the food has lowered concentrations below that detectable by the previous most sensitive method of detection.

Delay is an inevitable result of certain types of regulation, as in those areas requiring premarket approval. Delay also occurs administratively—e.g., when shortages of qualified personnel or turnover in personnel prevent prompt review of applications, or when a regulatory reviewer is unsure of what decisions to make and consults extensively within the agency before reaching a decision. Litigation over an agency's decisions and judicial review of these decisions impose further delays. These delays can be significant enough to affect the expected economic return on an innovation, which might cause the petitioning company to abandon the product and make investments elsewhere. Delay may, in effect, extend the life of already approved products, and, if costly, can impede the entry of small businesses into the particular market. Delay can also reduce the effective patent life of a new product, affecting its return on investment.

Regulation can also have other effects on innovation. It can affect the psychology of officials of private firms in conscious ways (e.g., when officials make decisions with an eye toward the likely reactions of the regulating agency) and unconscious ways (e.g., because officials have been accustomed to having to meet regulatory requirements). Furthermore, disclosure of data in support of an application for a new product approval can help another manufacturer compete with the original manufacturer.

REGULATION OF DRUGS AND MEDICAL DEVICES

The responsibility for Federal regulation of drugs and medical devices rests with FDA. FDA's regulatory modes are: 1) the establishing of standards, 2) the premarket notification process, 3) the premarket approval process, and 4) policing.

Policing typically occurs in lawsuits by FDA against violative products or firms. This mode is employed, for example, in the regulation of the labeling of medical devices. Establishing standards is a way of prescribing requirements for products or processes. For example, regulations governing "good laboratory practices" specify the mandatory, or in some cases the recommended, characteristics of the well-designed, properly conducted preclinical study. The premarket notification process gives FDA the opportunity to veto a firm's plans before they can be implemented. The 1976 Medical Device Amendments require that a firm intending to distribute a device for the first time notify FDA 90 days in advance to permit the agency to determine whether the device requires premarket testing and evaluation.

The premarket approval process is used by FDA to regulate drugs and certain devices. In the case of prescription drugs, a manufacturer must conduct tests for efficacy and safety on the drug, submit the data to FDA and obtain its approval before the drug can be marketed (244). FDA becomes officially involved in the development process for a new drug when its sponsor files a "notice of claimed investigational exemption for a new drug" (IND) for permission to test it in humans. There are three phases in the clinical investigation, and each phase must have been preceded by specified animal tests. (Animal test requirements for contraceptives are more stringent than the requirements set forth below for other drugs).

Phase I studies are investigations of a new drug's clinical pharmacology to determine levels of tolerance (toxicity), followed by early dose-ranging studies for safety (and, in some cases, efficacy) in selected patients. The total number of both healthy volunteers and patients, which varies with the drug, ranges from 20 to 50. If the drug is found to be safe, the manufacturer can proceed to the next phase of testing. Phase I studies must be preceded by 2- to 4-week studies in two animal species.

Phase II studies are designed to demonstrate effectiveness and relative safety of a new drug and are carried out on 100 to 200 patients under controlled conditions. If the drug's therapeutic value is demonstrated and there are no serious toxic effects, the manufacturer can proceed to the next phase. Phase II studies must be preceded by 90-day studies in two animal species.

Phase III studies are expanded controlled and uncontrolled clinical trials, involving 500 to 3,000 patients in usual medical care settings (clinics, private practice, hospitals). At least two well-controlled clinical trials, accompanied by complete case records for each patient, are usually required by FDA for approval of a "new drug application" (NDA).

If these clinical trials are successful, the drug's sponsor may file an NDA. An NDA is a request for FDA's permission to market the drug. Chronic animal toxicity studies (1-year dog, 18-month mouse, and 2-year rat studies) must be completed by the time of NDA submission. If the FDA review finds the effectiveness and toxicity data acceptable, the application is approved. Since 1962, FDA has reviewed over 13,500 applications for INDs and has approved about 1,000 NDAs (154).

The Medical Device Amendments of 1976 to the Food, Drug, and Cosmetic Act greatly expanded FDA's role in regulating medical devices. Prior to the 1976 amendments, FDA had classified devices such as soft contact lenses, pregnancy test kits, intrauterine devices, nylon sutures, and hemostats as "drugs" (359). The U.S. Supreme Court ruled in 1969 that this move was justified since Congress intended the public to be protected from unsafe and ineffective devices (299).

The Medical Device Amendments of 1976 established a three-tiered system of controls on medical devices. Class I devices are subject to general controls only; Class II devices must meet performance standards; and Class III devices must have premarket approval.

Class I devices are subject primarily to the Food, Drug, and Cosmetic Act's basic prohibition against misbranding and adulteration. Class I controls apply to accuracy in labeling and the sanitation and physical integrity of low-risk medical devices. All devices must meet these minimum standards. FDA also has the power to ban any device, regardless of classification, which presents a substantial deception or an unreasonable and substantial risk of illness or injury that is not correctable by labeling.

Class II controls are placed on devices for which general controls alone are judged insufficient, but about which sufficient information exists or could be developed to establish performance standards for the device. Under the 1976 amendments, existing voluntary standards could be used, but legal counsel advised that such actions would violate due process, as the "voluntary" standard might become essentially "mandatory" with the FDA stamp of approval, circumventing the opportunity for public comment and discussion (247).

Class III controls are comparable to the premarket approval process for drugs. These controls are applied when general controls or performance standards may not provide reasonable assurance of the safety and efficacy of a device which is life-sustaining, life-supporting, implanted, or presents a potential unreasonable risk of illness or injury, or when performance standards cannot be developed. Any device which was classified as a "drug" before the amendments is automatically assigned to Class III unless reclassified. Any device developed after the enactment of the amendments which is not judged by FDA to be "substantially equivalent" to a preamendment device in Class I or Class II will also be assigned to Class III and require a premarket approval application. In the first 4 years after implementation of the 1976 amendments, about 98 percent of the listed devices in the 10,540 premarket notifications received were declared "substantially equivalent" to a preamendment Class I or Class II device (260).

The Medical Device Amendments also allow FDA to permit developing and marketing approval of a Class III device under a "product development protocol," where FDA and the manufacturer agree in advance on a plan for the development, testing, and release of the device. This approach has not been implemented.

The 1976 amendments require any distributor of a device intended to be marketed for the first time to file a notice with FDA at least 90 days in advance to permit the agency to decide whether the device needs premarket approval to assure safety and efficacy. FDA permits earlier distribution if it concludes and notifies the distributor that premarket approval is not required. If the 90 days pass without comment from FDA, marketing can begin. In 1981, FDA estimated that 2,300 premarket notifications would be reviewed.

Industry often uses FDA approval to advantage in its marketing strategy. All results of clinical investigations will ultimately be included in a package insert, product data sheet, or physicians' brochure, which are FDA-approved generators of promotional claims (300).

MEDICAL AND SURGICAL PROCEDURES

Except insofar as State laws require that medical and surgical procedures be performed by physicians and that hospitals have certain facilities if they are to carry out certain procedures, medical and surgical procedures are essentially unregulated. State licensing statutes that define who can and cannot practice medicine (dentistry, etc.) preclude other technical personnel from performing many such procedures. Laws that restrict the performance of procedures to licensed facilities such as hospitals deflect from these settings innovative organizational arrangements such as home birth delivery and outpatient surgery.

Regulation of the practice of medicine is a State function carried out by State medical licensing boards.

However, State medical licensing boards primarily regulate entry into the practice of medicine and do little to monitor the continued competence of licensed physicians beyond assuring that they meet requirements for continuing medical education. However, a Federal program, the Professional Standards Review Organizations (PSROs), was enacted in 1972 to review medical care delivered to persons eligible for Medicare or Medicaid coverage. (As this program's functions relate more to Federal reimbursement for medical services, it will be described in the section below on reimbursement).

REGULATION OF CAPITAL INVESTMENTS

Three related Federal programs have been enacted in an attempt to regulate capital investment: 1) section 1122 review, 2) State certificate-of-need (CON) laws, and 3) the National Health Planning and Resources Development Act.

Since 1972, section 1122 of the Social Security Act has required the Medicare and Medicaid programs to withhold funding for depreciation, interest, and return on equity capital for certain investments found inconsistent with planning objectives by a health planning agency. The provision applies to investments of more than some specified amount (initially $100,000) and covers changes in beds and services that are provided by certain health care facilities, such as ambulatory surgical facilities. Health maintenance organizations (HMOs) are included, but private physicians' offices are explicitly exempted. In 1977, 37 States had contracted with the Department of Health and Human Services* to conduct section 1122 reviews.

The effect of section 1122 review is controversial. Since the statute excludes operating expenses and physicians' services, only a small percentage of a provider's total revenue may be at risk of scrutiny or control. For example, the operating expenses of computed tomography (CT) scanners account for as much as 50 to 75 percent of the technical expenses (279).

State CON laws, in effect, constitute a franchising process for potential adopters of expensive medical technologies. Enacted by 35 States by 1977, these laws require prior approval by the State of investments above a certain threshold (now usually $150,000 or more). Local health systems agencies have responsibility for areawide planning and initial CON review. Although the laws vary, most apply to hospitals and nursing homes. Like section 1122, most CON laws exempt private physicians. Sanctions include denial of operating licenses, court injunctions, and fines.

The National Health Planning and Resources Development Act of 1974 required States to pass CON laws by 1983 as a condition of future Federal funding under the Public Health Service Act, the Community Mental Health Centers Act, and the Alcohol Abuse and Alcoholism Act. The 1974 planning act generally applied to the same facilities covered by section 1122 review. However, the 1979 planning act amendments exempted HMOs from having to secure a CON for inpatient investments because of a belief in HMOs' efficiency.

In an early study of CON laws, Salkever and Bice (333,334) reported reduced hospital expenditures on beds, but unchanged overall hospital investment. Faced with greater control over beds, hospitals may have channeled their investments to other technologies. Furthermore, as Ginsburg (162) had found earlier, occupancy was positively associated with bed expansion, although occupancy rates had no apparent effect on total hospital investment.

Cromwell, et al. (75), investigated the effect of CON laws on the adoption of specific technologies. CON appeared to reduce adoption rates for expensive, widely adopted technologies—namely, X-rays and cobalt and radium therapies—but did not affect other technologies examined.

The existence of planning legislation was not correlated with interstate differences in the adoption of the CT scanner (392). In fact, impending legislation may have spurred adoption as providers rushed to place orders before the law applied to CT scanners. Such an effect may have occurred in California, whose 1976 law exempted equipment already ordered (12,19).

Reimbursement Policy

In contrast to regulation, which is often seen as having constraining effects, the growth in third-party coverage of medical care is seen as a major cause of the excessive adoption and use of many medical technologies (142,331). It is important to keep in mind, however, that just as regulation is but one influence on innovation, reimbursement policy is but one contributor to the overall tendency to adopt and use medical technologies at excessive levels. Other factors include competition among hospitals to achieve quality and prestige to attract patients and physicians, public demand for sophisticated technologies, increasing specialization within medicine, physicians' desires to do as much as possible for their patients, uncertainties related to what constitutes appropriate use, and the defensive overutilization of medical tests and procedures because of the threat of malpractice suits.

---

*Then the Department of Health, Education, and Welfare.

There are two basic forms of payment mechanisms in the U.S. medical care delivery system: cost-based and charge-based (306). Government programs, primarily Medicare and Medicaid, were developed to "buy into" what was then perceived as a market pricing system. When the statutes were enacted in 1965, the legislation established the principle that the Government purchaser would pay institutional providers the costs of services to patients. Physicians were to be paid their "usual, customary, and reasonable" fees. The assumption was that Government was buying at the margin and would not affect the average costs of the system.

The 1972 amendments to the Social Security Act reflected a growing understanding that purchases of medical services were sufficiently large to affect purchase price and costs. Consequently, limits were placed on the amount which would be paid by Medicare to both institutional providers and physicians. Rather than being related to efficiency, these cost limits reflected rates of increase in charges over time.

PSROs were enacted into law in 1972 and consist of areawide groupings of practicing physicians responsible for reviewing care delivered to persons eligible for Medicare or Medicaid coverage. They help assure that services provided and paid for by Federal beneficiary programs are medically necessary and of a quality that meets locally determined professional standards, and that they are provided at the most economical level consistent with quality of care. PSROs are separate, independent, nonprofit organizations located in a number of designated geographic areas of the country. They are physician-dominated organizations; upward of 50 percent of all practicing physicians in this country nominally belong to the PSRO in their area, although usually only a small fraction of these members participate regularly in PSRO activities.

For a variety of organizational and legislation reasons, PSROs have first concentrated on reviewing inpatient care provided in short-stay hospitals. One of their hospital-review activities is traditional utilization review intended to reduce unnecessary hospitalization. A second review activity is profile analysis, by which PSROs retrospectively review patient care data (aggregated by, for instance, provider or physician characteristics) to highlight patterns of care. Such analyses allow PSROs to identify problems in the use of services and to set objectives for changing the use of services. A third major type of hospital review activity is the "medical care evaluation" study which focuses more on quality of care than on cost containment.

Some PSROs, especially those with long experience in hospital utilization review, have moved beyond these activities to take on utilization or quality of care review in other facilities or medical settings. The major topics of such studies are ancillary services (virtually all services except for room and board, and nursing, dietary, or physician services in the hospital), long-term care review, and ambulatory care review. All three types of studies have been done in demonstration projects during the late 1970's and have been carried on since then by some PSROs, often as "special initiative" studies. At one time or another, as many as one-quarter to one-third of all PSROs had engaged in ancillary services or long-term care review; ambulatory care is, so far, a less well-developed field.

Several PSROs (or separately incorporated analogs) do utilization review for private firms on a contract basis. Perhaps as many as one-quarter of PSROs were engaged in such review as of 1980, and they covered patients whose care was financed by private insurance companies, self-insured corporations, the Civilian Health and Medical Program of the Uniformed Services, labor unions, and municipal governments.

In addition, several PSROs over the past few years have engaged in cooperative research projects, including projects related or analogous to technology assessment. One example of such a project is an ongoing RCT to evaluate different educational interventions intended to reduce the use of an outmoded obstetric practice (X-ray pelvimetry) in hospitals during deliveries. Other PSROs have collaborated in studies of variations of hospital use for several conditions such as myocardial infarction (heart attack) or gall bladder surgery.

Although the PSRO program pursues many objectives and tasks, the most visible have been those related to utilization and costs of hospital care. The 8 years of the program have not produced the desired reductions in hospital stays or, especially, in the costs of Federal health programs such as Medicare. Improvements in quality of care, although less well documented than the effects on costs, suggest that PSRO activities have ranged broadly across diagnoses and services. Currently, the Reagan administration is deemphasizing PSROs by defunding those thought to be ineffective and by consolidating areas.

There are two widely used mechanisms to set reimbursement levels in the "private" sector of the medical care market. One mechanism is the cost-based Blue Cross/Blue Shield reimbursement system. In many ways, this system is similar to the Medicare program. Hospitals are reimbursed the "reasonable" cost of providing care to patients, and physicians are paid "reasonable" fees. The second mechanism is payment for billed charges. This approach is used by some Blue Cross/Blue Shield plans and in all contracts established between patients and other insurers. Under this approach, all or some of the charges of hospitals and other medical providers are paid through insurers, unless there are copayments and deductibles which are paid by the patient. There are also patients not covered by Government or other insurers who are responsible for their own bills. Billed charges are more like a market mechanism, except that demand is not directly affected by the income or wealth of the patient. A third payment mechanism, not yet very widespread, is capitation, whereby a fixed amount is paid for each patient per time period, regardless of the health services provided. The capitation method generally involves the integration of financing and services, thus placing the provider of care at financial risk.

INFLUENCE OF REIMBURSEMENT ON THE DEVELOPMENT, ADOPTION, AND DIFFUSION OF MEDICAL TECHNOLOGY

When coverage has been offered from the outset for new and experimental medical and surgical procedures, a high level of reimbursement has been justified on the basis of the special skills and large amount of professional time required, and perhaps on the basis of increased risk. But, when such procedures have become routine, requiring less time and skill and posing lesser risks, fees for the procedure have usually increased rather than fallen (316).

Several examples have been provided by Blue Shield of California (40). Phakoemulsification of the crystalline lens, introduced as an alternative to lens extraction for cataract, is—once learned—shorter and no more complex than standard lens extraction, yet surgeons initially attempted to charge 25 to 30 percent more for the new procedure than they charged for the older one. The Blue Shield Medical Policy Committee disallowed the increase. Another example is the flexible fiberoptic endoscope. This new instrument is easier to use than the standard rigid instrument, yet physicians introducing the new procedure attempted to charge 25 percent more. Similarly, orthopedic surgeons who introduced arthroscopic menisectomy for torn knee cartilage wished to charge the full fee for the standard open arthrotomy and an additional fee for arthroscopy. In this instance, Blue Shield of California agreed to pay the full arthrotomy fee and an additional 50 pecent of the arthroscopy fee. The rationale for Blue Shield's concession was that carrying out the simpler procedure might eliminate the need for many days of hospitalization and laboratory tests, with a considerable net savings in total charges.

Allowing a simpler procedure to be billed as a more complex procedure results in questionable increases in physicians' fees. In the example just cited, the large difference in allowable charges when an operative procedure is added to a diagnostic procedure offers a strong invitation to remove some tissue during arthroscopy. During the diagnostic examination of the knee, a small piece of redundant synovial membrane may be seen—a finding of no great import. Removing a piece of this tissue makes the procedure a "synovectomy," for which the customary charge is $1,300, rather than simply a diagnostic arthroscopy, for which the customary charge is $500. The above scenario presents a situation that may be reasonably justified medically, but, even interpreted generously, there is a clear fiscal invitation to perform a procedure that is more, rather than less, complex.

There also is a much more serious consequence of the manner in which charges are submitted for experimental procedures. With increasing scrutiny by third-party payers of bills submitted for new procedures and with more than occasional denial of payment for such bills, there is a strong incentive for physicians to request payment for a standard procedure rather than the new one. This is also encouraged by the fact that new procedures often do not have a procedure code number, by which most bills are processed. Requesting payment for a standard procedure may simply reflect an honest effort to use whatever code number seems most nearly to approximate the procedure actually performed. Whatever the motives, the net result is that the identity of the new procedure may be concealed, and the fact that an experiment has been carried out may not emerge.

In bills submitted to Blue Shield of California, there is an approximately 15-percent error rate in the coding of all procedures (39). It is estimated by the medical director that 1 percent of the errors involve the use of existing codes for procedures to which new codes have not been assigned.

Because it is difficult to define exactly what constitutes "accepted medical practice," the new procedures that have the best chance of being reimbursed are the ones which deviate the least from existing procedures which are already being reimbursed. The Federal Government, for example, has traditionally favored coverage of new technologies perceived to be modifications of existing interventions (270). The incentives, therefore, are toward the development of parallel procedures or extensions of existing technologies.

For procedures that deviate substantially from accepted medical practice, the reimbursement system may require considerable testing for safety, efficacy, and costs to determine if they offer sufficient contributions to compensate for their deviation from standard medical practice. These circumstances have several implications. First, when procedures remain outside the coverage range, they may also suffer the fate of anonymity, neglect, lack of funding, or underutilization.

An obvious example is the traditional exclusion from most insurance plans of much preventive medical care, most notably screening services. Second, the scrutiny of radical innovations rather than of incremental improvements may be misplaced to the extent that the growth in medical expenditures is the primary reason for such scrutiny. The collective expense of small tests and procedures is arguably far greater than that of a few "big ticket" technologies (249). Third, if radical innovations have the most difficulty in receiving favorable coverage decisions, innovators might be inclined to pursue less radical but more easily accepted innovations. This is a difficult hypothesis to test, as radical innovations have less chance of commercial success than minor innovations; but once they penetrate the market, the magnitude of their commercial success is greater than for minor innovations. Fourth, as discussed above, a technology-by-technology approach to coverage decisions, with priorities determined by how radically each technology differs from existing ones, may lead those seeking payment for the use of new technologies to submit their claims for payment under the guise of accepted procedures.

Under either cost reimbursement or charge payment, third-party payments generally are intended to cover the full costs of new technologies, including purchase, maintenance, or operation of equipment; the leasing of equipment; the cost of drugs; or the facilities and equipment needed for a procedure (19). One would expect that greater adoption of technologies would occur under these relatively price-independent conditions than would occur under a more price-sensitive system. Cromwell, et al.'s, interstate analysis (75) found that the percentage of revenues from third parties significantly and positively related to a hospital's adoption of expensive technology. Russell (331) found that adoption of cobalt therapy and electroencephalography occurred faster when the level of insurance coverage was higher and proceeded more rapidly as that level grew. She also found that a greater contribution to hospital costs by Medicare was associated with increased adoption of cobalt therapy, intensive care beds, and diagnostic radioisotopes. And Willems (392) concluded that open-heart surgery spread more quickly in areas with faster growth in insurance coverage.

Third-party reimbursement can also indirectly affect the adoption of technology by changing the availability of financial capital to potential adopters. A prominent example is the Medicare program, which reimburses institutional providers for capital as well as operating costs. Medicare payment for allowable capital costs such as depreciation and interest provides a source of internally generated funds (28). Third-party coverage, especially by Medicare and Medicaid, has also reduced hospitals' risks of bad debts, thereby improving their standing as credit risks to private lenders. Other changes in governmental programs, such as the Hill-Burton program for funding medical facility construction and modernization, as well as various tax-exempt bond programs, have affected the source of financial capital.

In addition to affecting the adoption of technologies, the extent of third-party coverage would be expected to affect the use of technologies. Data on the use of specific technologies are generally lacking, however. Cromwell, et al. (75), found that many hospital technologies are underutilized after being adopted. Nonprofit hospitals in the Boston area were using automated analyzers, patient monitors, and, in teaching hospitals, diagnostic X-rays, at only about half of capacity. Willems (392) considers such underuse as presumptive evidence of the hospitals' overinvestment in new equipment.

It is not clear how this relatively price-independent adoption of medical technologies is used by medical care providers to compete with one another. As summarized by Banta, et al. (19):

> Studies of hospitals have found no definite relationship between measures of competition and adoption. The situation is complex, because the characteristics of the market may relate not only to competitiveness, but also to the availability and sharing of information and to local standards of practice. The evidence conflicts, depending on the characteristic used and the technology studied. Russell (331) found that concentration of market power among a few large hospitals did not appear to influence the adoption of three common and two prestige technologies, but that hospitals in more concentrated markets were less likely to adopt open-heart surgery. Prior adoption in a locality reportedly speeded the adoption of intensive care units and electroencephalographs, but not diagnostic radioisotopes, open-heart surgery, renal dialysis, cobalt therapy, and computers (75,331). In urban areas, greater adoption of radioisotopes and electronic data processing occurred where there were many hospitals per capita, the hospitals were of similar size, and they were close to other hospitals (212,301).
>
> Different patterns have also been observed between adoption and the number of physicians per capita. Facing a low physician-population ratio, hospitals may compete for physicians through technology adoption. On the other hand, fewer physicians may exert less pressure for adoption. The adoption of CT scanners and radioisotopes appeared unrelated to the physician-population ratio (301,392). However, greater adoption of intensive care units, open-heart surgery, cobalt therapy, and renal dialysis occurred among States with higher ratios (75).

Thus, even though current payment mechanisms for medical care services can lead to excessive adoption of medical technologies, there are still constraining factors which make it clear that cost is not the only factor which influences adoption.

Discussion and Conclusions

No one factor seems to distinguish successful from unsuccessful innovations (252). The key events identified in studies of successful innovations depend on the choices of innovations for study. Failure to find a cause-and-effect relationship between one variable and success in innovations is not surprising, however, given the multiple factors in the innovation process.

It is therefore also not surprising to find that the impacts of regulatory and medical care reimbursement policies on the innovation process are difficult to separate from the impacts of other factors. In the regulation of drugs, for example, the evidence points toward regulation as contributing to, but not as being the sole or primary determinant of, higher R&D costs, greater concentration of new drug development in fewer and larger firms, and an orientation toward the epidemiologically and commercially more important diseases.

The impacts of regulatory policies are not well understood. The availability of data on R&D costs, number and size of firms producing new products, the number of new products, etc., almost compel researchers to focus on these parameters in their evaluative work. Quantitative rather than qualitative analyses are what most people expect in order to translate complex relationships into simple terms such as "bottom line" numerical estimates of how regulation affects the innovation process. The result is a focus on easily identifiable costs and a neglect of difficult-to-quantify benefits, and most of the controversy centers on whether these identified costs are due to regulation or other factors such as, for example, existing trends in the drug industry at the time of the 1962 drug amendments.

On the other hand, the studies that focus on costs might be thought of as providing presumptive evidence of the costs of regulation, thereby shifting the burden of proof to regulation's advocates to counteract the evidence with findings on the benefits of regulations. The problem with this approach is not only that health impacts cannot be measured adequately, but that even if they could be, there are no unambiguous methods to compare the costs and benefits (270).

Regulation is meant to alter the market forces controlling the innovation process, so it should come as no surprise that observable economic measures are altered. The fundamental question underlying the debate over the absolute or net costs (where benefits are considered) is whether the achievement of the social purposes of regulation is worth the costs. With respect to drugs, even when evaluation of the impact of regulation focuses on absolute costs, most critics of the current regulatory process call for marginal alterations, not radical changes. Such changes include more flexibility in efficacy and safety testing, speeding up of the premarket approval process, and the use of postmarketing surveillance systems.

For regulation of medical devices, there seems to be no major opposition to the law per se, only a wait-and-see attitude and differences of opinion as to how FDA is implementing certain provisions of the Medical Device Amendments of 1976. For example, FDA has proposed six broad categories of devices which would be subject to premarket testing: 1) invasive devices intended to pierce the skin or mucous membranes; 2) implantable or prosthetic devices; 3) energy-introducing devices; 4) medicinal gas devices; 5) devices, other than in vitro diagnostic products, that are intended for use in diagnosing of disease or monitoring physiological functions; and 6) in vitro diagnostic products intended to provide information which will be used, interpreted, or analyzed by a health professional.

Industry associations, such as the Health Industry Manufacturers Association, urge continued case-by-case determinations and are opposed to these broad categories, because they believe that there is no demonstrated need for the rule and point out that FDA has not conducted cost-impact studies (176).

Much as an IND does, "investigational device exemption" (IDE) regulations describe the requirements for clinical investigation and the responsibilities of the manufacturer, clinical investigator, and the institutional review board. None of this documentation was previously required, and industry has protested that the mandated process would simply give rise to additional costs and delays as well as automatically trigger an FDA inspection of facilities, if one had not been previously done, when an IDE was submitted. Therefore, FDA has been requiring IDEs only for devices which require premarket approval (Class III devices) (359).

A proposed "mandatory experience reporting" rule would require manufacturers, distributors, and importers to report to FDA any device which may have caused injury or death, has a deficiency that could result in death or injury or give inaccurate diagnostic information, or is the subject of remedial action. The proposed rule would require those covered to report device-related deaths within 72 hours after receiving a complaint, injuries within 7 working days, and remedial action or communication with distributors, health care practitioners, or users within 2 working days (8). In response to FDA estimates of $20 per report, the Health Industry Manufacturers Association estimates that the entire industry would incur a total annual cost in excess of $400 million (177).

The wide availability of medical insurance contributes to overadoption of many new technologies, but other factors in the medical care delivery system have significant influences. Some of these other factors can add incentives to overutilize new technologies, but additional factors seem to be keeping the rate of adoption of new technologies below the level expected if costs were the only or primary criterion influencing adoption. Thus, just as regulation has contributed to existing trends in the drug industry, current reimbursement policies also have contributed to overadoption of new technologies but cannot be credited as the determining factor.

In addition to its contribution toward excessive demand for new technologies, current reimbursement policy has another significant effect on the innovation process. Radical innovations, which by their very definition often fall outside generally accepted medical practice, tend not to be reimbursed and thus may be less likely to be developed. The current system discourages the identification of new procedures as such.

This appendix has described the effects of regulation and reimbursement policies on the innovation process as separate issues, but they clearly are interrelated. As new medical procedures develop, they often make use of new drugs and devices or use existing ones in modified ways. In either case, the drugs and devices generally have to pass through the regulatory process. Until they are approved, regulatory review acts as a constraint on the adoption and dissemination of the procedures in which they are used.

Regulatory review is generally limited to the technical questions of safety and efficacy, without consideration of the costs or relative values of the proposed drug or device once it reaches the market. In some cases, however, it goes beyond these questions. For example, in reviewing the injectable contraceptive Depo Provera, FDA used marketing as well as safety and efficacy criteria to deny approval. In that case, FDA denied approval not only because of its concerns over Depo Provera's cancer-causing potential, but partly on the basis that the patient population originally targeted for Depo Provera had diminished substantially as other methods of contraception and sterilization had become increasingly available and accepted (193). Nevertheless, the more usual circumstance is such as that found in the approval of the catheter used in percutaneous transluminal angioplasty (PTCA). In that instance, FDA released the catheter from investigational device status and approved its marketing for PTCA while the procedure itself was still considered by many to be experimental. Thus, while regulation of the accessories (i.e., drugs and devices) acts as a constraint on the adoption of the medical and surgical procedures in which they are used, once these accessories are released into the marketplace, they can act to stimulate use of procedures which are still experimental and not accepted medical practice.

Does this observation point to a strategy for medical technology assessment in which the criteria are similar for both regulatory and reimbursement purposes? From the review of the regulatory process, it appears that the system for regulation of drugs and devices meets certain social goals. Although economic considerations are important, these considerations point toward specifying how the present regulatory process can be improved, but not toward the infusion of economic measures into the regulatory criteria themselves.

The infusion of economic measures into the regulatory criteria themselves may be arbitrary and counterproductive. Users of innovations are important contributors both in determining the full extent of an innovation and in developing new innovations as spinoffs the exact results of which can never be determined beforehand.

The current regulatory process for medical technologies may need marginal changes, but the consensus seems to be that its social usefulness is worth the costs which it places on an innovation process dominated by a market approach. This conclusion is compatible with the common sense notion that society should focus on the use of the tools and not on the tools themselves to keep the constraints on the innovation process at a minimum while also addressing the issues of cost, quality, and appropriateness of medical care.

Current reimbursement policies both stimulate and constrain the development of new medical technologies. Possible modifications of these policies might well be examined for their potential.

Appendix E.—Case Studies of Medical Technologies

Introduction

The five case studies in this appendix are included as illustrative examples of the innovation process, including the development and adoption, of selected medical technologies. Of particular note are the effects which Federal research funding and regulatory and reimbursement policies have on the innovation process.

The case study on percutaneous transluminal coronary angioplasty (PTCA) was written by David Sawi, and that on hemodialysis and kidney transplant surgery by Katherine Jones. The other case studies on gastric freezing for the treatment of ulcers, hemodialysis for the treatment of schizophrenia, and maternal serum alpha-fetoprotein (MSAFP) were prepared by OTA staff.

Gastric Freezing

In the mid-1950's, a surgical leader in the United States, Owen Wangensteen (of the University of Minnesota Medical School), observed that iced saline solution lavaged into the stomach slowed gastrointestinal bleeding and that animal experiments showed reduction in stomach acid output following gastric cooling. From these observations, Wangensteen conceived of the idea of using gastric cooling for treatment of peptic ulcer disease (143).

In collaboration with a small refrigeration company, he developed a device to circulate alcohol at −15° C through a nasogastric tube to a balloon inserted into the stomach. After testing the device in dogs, he first tried it on one patient, then a dozen others (143).

In 1962, he reported his results in the *Journal of the American Medical Association:* no serious side effects, markedly reduced stomach acid output, immediate relief of ulcer pain, and radiographic evidence of ulcers healing (381). The 1962 report was extensively covered in the popular media. Wangensteen also sought to spread his procedure through professional meetings and publications, and the American College of Surgeons prepared an instructional film on the technique (143).

By the end of 1963, 1,000 devices had been sold and 10,000 to 15,000 procedures had been performed nationwide (143).

Beginning in 1963, however, the efficacy and safety of gastric freezing began to be questioned. In 1964, there began to appear published reports concluding that acid suppression was limited or unrelated to pain relief, symptomatic improvement was short-lived or due to placebo effects, and important risks were present. Variations in the technique were used and became arguing points for valid use, but by 1966 the technique was rarely used. Furthermore, Wangensteen lost his support from the National Institutes of Health (NIH) for research in gastric freezing, although he still thought the procedure worthwhile and did not believe his earlier reports to be inaccurate (143).

Hemodialysis for Treatment of Schizophrenia

Schizophrenia, a disorder characterized by misinterpretation and retreat from reality, delusions, hallucinations, ambivalence, inappropriate affect, and withdrawn, bizarre, or regressive behavior, is estimated to afflict 2 to 3 percent of the population. Because of its relatively high prevalence rate, onset in adolescence and early adulthood, and lifelong chronicity, this disease imposes a major toll of disability and a great economic burden on our society. Although effective treatment for the disease could help large numbers of people, ineffective new interventions pose the prospective risk of wasting large sums of money. Schizophrenia's unpredictable course in individual patients, coupled with the difficulty of assessing the health status of patients, objectively enhances the risk that ineffective or unproved modes of treatment will be adopted, perhaps even widely, in the management of schizophrenic patients (50).

In 1977, Wagemaker and Cade (379) created intense interest by claiming, in a report published in the *American Journal of Psychiatry,* dramatic improvement in five physically healthy schizophrenic patients treated with weekly dialyses for up to 16 weeks. Since then, Wagemaker and Cade have continued this treatment. Wagemaker has dialyzed an additional 15 patients, reporting 100 percent success in 7 women and 3 good successes, 3 partial successes, and 2 failures in 8 men (109).

Scattered reports in the literature between 1925 and 1960 had claimed improvement in schizophrenic patients given blood transfusions either from remitted schizophrenics or from healthy volunteers, but Feer, et al. (139), apparently were the first to use dialysis. These investigators reported in 1960 that three out of five schizophrenics improved after only one or two hemodialyses (139).

There is some evidence that hemodialysis may remove a circulatory psychotogen, as Palmour and Ervin (285) reported a substance characterized as a beta-endorphin in the dialysate of the patients treated by

Wagemaker and Cade. A 100-fold decrease in the substance, corresponding to improved clinical functioning, was found in 14 of 16 treated patients. This claim has not been duplicated and reported by others.

Because of the positive report by Wagemaker and Cade, the Clinical Research Branch of the National Institute of Mental Health (NIMH) funded three research projects using a double-blind design. Dr. Wagemaker, of the University of Louisville (Kentucky), developed the first project beginning in September 1978 through his "own initiative with considerable institute staff consultation to develop an acceptable protocol." This study is near completion but has not been reported. The other two projects are being conducted at the University of Maryland (Baltimore) and the University of Washington (Seattle), both starting in September 1979. In addition, the NIMH Intramural Program conducted a small study, the results of which were published in *Science* in March 1981 (338). Of eight chronic schizophrenics, none of the patients improved during active dialysis, and four patients worsened.

In September 1980, the National Center for Health Care Technology (NCHCT) issued a memorandum at the request of the Health Care Financing Administration (HCFA) for a recommendation regarding the use of hemodialysis in the treatment of schizophrenia. NCHCT concluded that the evidence on its safety and efficacy was inconclusive and recommended that the procedure not be covered under Medicare (109).

NCHCT later commissioned an analysis to estimate the economic effect of the decision not to reimburse for the procedure. The study found that under the central estimate, by which 1 percent (4,000 patients) of Medicare-age schizophrenics would receive dialysis in 1984, annual costs would reach a peak of $15.4 million in 1983, and the present value of total costs over a 7-year life of this practice would total $29.3 million. High and low assumptions, under which a minimum of 0.25 percent (1,000 patients) or 5 percent (20,000 patients) of the pool of schizophrenics are dialyzed, result in peak annual costs of $3.9 million to $82.1 million and total costs of $7.7 million to $149.9 million.

If similar fractions of the entire schizophrenic population were to be treated, the total costs of treating schizophrenia with dialysis in this larger pool would be five times higher, or $39 million to $750 million. While funds for patients not treated under Medicare would be provided through Blue Cross/Blue Shield, insurance companies, and private individuals, a portion of this cost would be assumed by the Federal Government as diminished tax revenues due to increased medical expenditures (50).

This case study characterizes NCHCT's former role in responding to HCFA's Medicare coverage issues. It also suggests the utility and need for a continued systematic coverage evaluation process within the Public Health Service. Expenditures for dialysis in the Medicare population alone could have created substantial cost burdens in the absence of a clear Federal policy. In view of the unproved effectiveness of the treatment, NCHCT's recommendation was based on medical/scientific grounds. As this case illustrates, however, it is very difficult to separate cost implications from such decisions.

Percutaneous Transluminal Coronary Angioplasty*

by David Sawi, M.B.A.
Kaiser Foundation Health Plan, Oakland, Calif.

Introduction

On January 16, 1964, Charles T. Dotter and M. P. Judkins performed the first percutaneous transluminal angioplasty (PTA) (6,27). The patient, an 83-year-old woman, was referred to the University of Oregon Medical School Hospital with a disorder of the left leg. She had a 6-month history of pain and infection of the left foot and toes, and a gangrenous appearance in three toes had occurred within the previous 3 months. Angiographic examination revealed a 0.5 cm long atherosclerotic obstruction of the left superficial femoral artery at the level of the adductor hiatus. The patient was considered unsuitable for vascular surgery owing to her age, poor cardiac condition, and bad runoff. Because of her advanced gangrene, low thigh amputation was advised, which the patient refused.

It was then decided to attempt catheter dilation. Treatment, using a coaxial double catheter, lasted a short time and gave excellent results. The patient's pain disappeared within hours, and the patient was ambulatory within weeks. Repeated followup angigraphy confirmed the patency of the treated artery. The patient continued to walk without difficulty up to her death at the age of 86 years.

With the increase in lifespan over the past few decades, more patients now need surgical vascular reconstruction. This need increased the demand for more centers specializing in vascular surgery and equipped with intensive care facilities (30). Additionally, surgical measures were not always successful when applied to smaller arteries and faced limiting factors such as technical difficulty and operative trauma (6).

---

*NOTE: Reference citations for David Sawi's case study on PTCA appear on p. 174.

Description of Procedure

PTA is the noninvasive, mechanical treatment of vascular obstructions with the use of catheters (5). The procedure is done by one of two methods, depending on the nature of the lesion.

The first method, called transluminal dilation, is correcting a stenotic lesion by enlarging the diameter in the constricted lumen (or passage). For example, in Dotter's initial procedure (6), a tapered, radiopaque, Teflon dilating catheter of approximately 0.1-inch outer diameter was slipped over a coil-spring catheter guide of about 0.05-inch outer diameter. The catheter guide had been passed down the lumen until its tip had traversed the stenosis. The passage of the dilating catheter over the catheter guide enlarged the stenotic lesion by exerting outward pressure on the lumen. This method was refined by the introduction in 1974 of a catheter with a distensible (balloon) tip, which, when inflated, exerted outward pressure on the lumen (12).

The second method, called transluminal recanalization, is used to correct a vascular occlusion by creating an artificial lumen through an occluded segment. The catheter guide is passed through the occluded segment, and then the passage is enlarged by the introduction of the dilating catheter over the catheter guide (27). This approach is possible because the atheroma causing the obstruction consists of a low-density fatty material which has the characteristic of inelastic compressibility (12,30).

The above procedures are often done in conjunction with anticoagulants and platelet aggregation inhibitors, which seem to improve outcome (30).

There are basically two classes of transluminal angioplasty: peripheral angioplasty (PTA) and coronary angioplasty (PTCA). The technique performed by Dotter and Judkins in 1964 (6) was peripheral angioplasty. In September 1977, Andreas Grüntzig performed the first nonoperative transluminal angioplasty of coronary arteries in a human being (12).

Indications and Alternatives

Indications for PTA are disabling claudication (limping, cramp-like pain due to inadequate blood supply), salvage effort prior to amputation, short stenosis in a large caliber, accessible artery, and little likelihood of success with reconstructive surgery (31).

Indications for PTCA are more restrictive. The patient should have a short history of anginal pain and should be experiencing disabling angina. The obstruction should be within a single vessel to minimize risk should complications occur. Grüntzig estimates that of patients with coronary heart disease, 3 to 5 percent of the older medical population and 10 to 15 percent of the younger population are suitable for PTCA (12).

Alternatives to PTA appear varied. At times, PTA is done as an alternative to surgery, whereas in other instances, it is done when surgery is contraindicated due to high risk and little chance of success (i.e., in elderly patients with prior heart problems) (5). Also, PTA can be an alternative to amputation.

PTCA is an alternative to coronary artery bypass graft (CABG) surgery, but only for a small percentage of patients. As of 1979, it appeared that the 5-year survival rate for PTCA was comparable to that of bypass surgery. However, improved blood flow to ischemic segments was also noted (23).

Historical Development

The development of PTA can be viewed as having four stages: 1) the technical development leading to Dotter and Judkin's initial peripheral angioplasty, 2) the period thereafter during which clinical tests reaffirmed the efficacy of the procedure and began to establish clear-cut indications, 3) the modification of the method by different types of catheters for the dilation technique, which lowered the rate of complications, and 4) the extension of the technique to coronary angioplasty (PTCA).

The technological development leading up to Dotter's initial procedure primarily related to the refinement of the catheter. Seidlinger (Sweden) first introduced the flexible catheter in 1953 (25). Meanwhile, Dotter (U.S.) had been refining a catheter for use in occlusion angiography (1958) and diagnostic angiography (1951). In 1962, at about the time Dotter commenced post mortem investigations using a coaxial dilation catheter, Nordenström introduced five types of balloon catheters for percutaneous insertion using a modified Seidlinger technique (22). Prior to Nordenström, balloon catheters had been introduced into the vessel via an incision. In February 1963, Fogarty (England) used a balloon catheter for the extraction of distal thrombosis (8). In the early 1970's, Postmann's (Germany) caged balloon catheter and Wholey's (U.S.) balloon catheters preceded Grüntzig's (Sweden) development of a viable and effective balloon-tipped catheter in 1974. It appears as though subsequent development of the catheter instrument has been primarily a refinement of Grüntzig's model.

It is clear from this brief overview that researchers in a number of different countries were essentially working on the same problem simultaneously. Two primary dynamics in the catheter development were the desire to minimize the possibility of embolism as a result of PTA and the desire to improve the patency rate on followup studies. In regard to this second point, Zeitler reports that between 1968 and 1980, he conducted a randomized study of 1,217 procedures. Treatment conditions included different types of catheters and different drug regimens (29). Zeitler found the patency rate of a newer, improved single Teflon catheter and the Grüntzig balloon catheter to be three times that of the coaxial dilating set.

Numerous other clinical trials have been conducted in addition to Zeitler's study (see table E-1). Two points should be made regarding these studies. One is that the outcomes seem to substantiate the claim that PTA is a viable alternative to incisive treatment and, in fact, can succeed where surgery might fail. The second point is that the trials in table E-1 are not, strictly speaking, comparable with each other. A number of flaws prevent this comparison:

- In later years, better patient selection increased the probability of successful treatment.
- The catheters were constantly being improved, especially in Grüntzig's balloon catheter.
- Criteria for a "successful outcome" are not consistent across trials.
- The skills of the physicians differed.

Additionally, the trials were not *randomized* clinical trials, in which patients are randomly assigned to treatment/no treatment conditions, or to treatment A (PTA)/treatment B (e.g., surgery) conditions. The relative efficacy of PTA was determined by comparing treatment outcome v. historical data on outcome using alternative methods. In spite of these studies' methodological flaws mentioned above, their results were sufficiently positive to encourage continued research.

Grüntzig's balloon catheter (1974) is generally credited with being the key to PTA's rapid diffusion (3,4). Grüntzig received research support to develop the balloon catheter from a European firm named Snyder. This seems to have been a significant variable. Dotter had been attempting to refine the balloon catheter for a number of years prior to Grüntzig's success. Dotter's lack of funding support hampered his efforts (4).

Notwithstanding a decision by Medicare to discontinue reimbursement of PTA, substantial sales growth (in units) is expected for the next few years (see fig. E-1). Based on figures provided by Cook, Inc., the size of the PTA catheter market for 1981 was approximately 100,000 units. This was expected to increase to near-

Table E-1.—Reported Clinical Trials of Peripheral Transluminal Angioplasty

Publication date	Trial dates	Observations[a]	+ %[b]	Comments[c]
1964	1964	15	>50%	Dotter, and Judkins: F,P,I
1966	1/64-9/65	113	~50+%	Dotter, et al.
1968	*	153	71%	Dotter, et al.: F
1969	*	59	64% (WA)[d]	Brahme, et al. (Sweoen): F
1971	*	161	70%	Zeitler (Germany): F,P,I
1973	*	100	71-89%	Wierny, et al.: F
1973	*	237	80-90%	Dotter: F
1973	*	25	84%	Grüntzig (Switzerland):
1974	*	43	81%	Dotter: I, with ballon catheter
1975	*	210	78%	Zeitler: I, with balloon catheter
1975	1969-75	534	74% (WA)[d]	Zeitler, et al.:F,P
1978	*	69	*	Zeitler: randomized catheter
1978	9/77-8/78	61	51%	Grüntzig: PTCA
1978	11/68-12/73	138	75% (50%)	Schoop, et al.: I
1978	*	1,184	74%	German study with
		206	92%	numerous contributors: F,I
1979	3/78-4/79	64	*	
	1971-3/78	188	86% (balloon)	Grüntzig and Kumpe
		48	64% (coax)	
		300		
1979	*	43	96%	Katzen: balloon catheter
1980	1978-80	172	80%	Colapin, et al. (Canada)
1980	1968-80	1,217	*	Zeitler

Key: *Unstated or unclear.

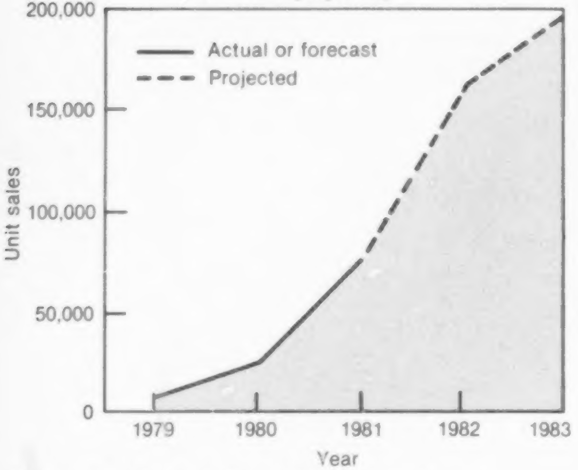
[a] Indicates number of procedures performed. Later reports undoubtedly include outcomes of early studies.

[b] Generally stated as patency rate within 2 weeks of procedure.

[c] F = femoral; P = popliteal; I = iliac.

[d] WA indicates weighted average primary success rate across various trial conditions.

Figure E-1 —Cook, Inc., Annual PTA Catheter Sales (actual and projected)



Cook, Inc.: 65 to 75 percent market share, $70,000 per unit

SOURCE: M. Kanne, Cook, Inc., Bloomington, Ind.

ly 300,000 units by 1983. After that, growth should level off to an annual rate of 10 to 15 percent (17).

Innovation of PTCA

In 1977, Grüntzig performed the first PTCA (see table E-2). There appear to be two major factors leading to the innovation of PTCA:

- PTCA is, essentially, an extension of the same concept, methodology, and technology as that used in PTA; to extend the concept to coronary arteries was nearly inevitable.
- Grüntzig's refinement of the balloon catheter provided the technology necessary to safely perform the procedure.

Diffusion of PTCA

At present, three factors are driving the diffusion of PTCA. The first factor is Grüntzig himself. Grüntzig's impact has been felt through the instruments he developed to perform the technique; his contribution to the literature (see table E- 2); his refinement of the procedure, with meticulous recording of numerous clinical trials; and the training he has provided to other physicians, both in Zurich and other sites to supervise initial operations.

A second factor is the Food and Drug Administration's (FDA's) decision in late 1980 to release the PTCA catheter device from investigational device exemption (IDE) status (19). An IDE status limits a company's ability to market a product by placing strict guidelines on the product's distribution and use. Each purchasing institution has to be designated as an investigator by FDA, and this involves considerable documentation and takes 2 to 3 months for approval. With removal of IDE status, the restrictions are removed, and marketing and purchasing of the catheter are simplified. USCI, one of the two manufacturers, claims that since its catheter received premarketing approval status, the number of centers which have ordered equipment to perform PTCA has doubled (9).

The third factor is the Interim Registry of PTCA at the National Heart, Lung, and Blood Institute. This registry receives voluntary information from centers performing PTCA within the United States and reports on successes, measurements, complications, and followups. The availability from the Interim Registry of information showing outcomes which are consistently comparable with alternative, invasive procedures nevertheless has a powerful influence in predisposing physicians to accept PTCA. It also creates pressures on third-party reimbursers to extend coverage to a procedure whose efficacy does not appear to be in doubt.

Table E-2.—Reported Clinical Trials of PTCA

Publication date	Trial dates	Observations	+ %	Comments
1977	—	—	—	Grüntzig: animals and post mortem studies
1978	—	—	—	3 post mortem and 1 animal study reported[a]
1978	9/77-12/77	7	86%	Grüntzig
1978	1977-78	29	70%	Grüntzig, et al.
1979	*	65	60%	Grüntzig, et al. United States and Switzerland, 5 centers
1979	1/78-7/79	50	68%	Grüntzig, et al.
1979	9/77-10/79	163	62%	Grüntzig

Key: *Unstated or unclear.
[a] *Circulation* 57 (suppl. 2):80, 1978

Considerable uncertainty exists over the completeness of the registry's data base. The cumulative number of procedures reported to the registry is shown in table E-3. As of May 1981, the registry had 100 centers reporting to it (19); however, USCI and Advanced Catheter Systems (ACS), the two manufacturers of the catheters used to perform PTCA, estimate the number of purchasers (i.e., centers) to be between 300 and 400. FDA's release of the catheter from IDE status in late 1980 made it available to users other than reporting centers. According to the manufacturers, therefore, the registry's data undercount the number of procedures actually performed.

Third-Party Reimbursement

It was initially thought that PTCA was not reimbursed by third-party carriers. This is widely believed by numerous professionals in the field. However, subsequent information indicated that this was not the case and that the situation is as follows.

First, Medicare, MediCal, Blue Cross, and Blue Shield do not reimburse for PTCA. The procedure is viewed as being investigational, and the primary reservation is the lack of data on clinical effectiveness. Data coming out of the Interim Registry and FDA's decision are viewed as important precursors to recognition from the aforementioned carriers (7,20).

Second, grants and free service have covered the costs in some institutions. For example, at Stanford University Hospital, PTCA has been recognized as a "research procedure." Thus, either research grants have covered the costs, or the service has been provided free (1). In 1981, PTCA was being reviewed to determine whether it could be called a "billable procedure," in which case third-party reimbursement would be sought to the extent that it is available.

Third, most private carriers* reimburse for PTCA to some extent, either knowingly or otherwise. In some cases, the charges for PTCA are "buried" in other, chargeable, items, such as catheter laboratory charges, coronary angiograms, etc. (7,19,26). More often, it seems, the procedure is openly identified as PTCA and charged accordingly.

Table E-3.—Reported Use of PTCA

Date	Cumulative number of procedures reported	Number of centers
June 1979	61	5
December 1979	200	—
June 1980		60
Fall 1980—FDA released catheter from list of investigational devices		
October 1980	504	—
February 1981	1,016	—
May 1981	1,800	100

SOURCE: S. Mullin, Interim Registry of PTCA, Cardiac Disease Branch, National Heart, Lung, and Blood Institute.

The latter is the procedure used, for example, by Steven Myler of San Francisco, who has an active practice in PTCA.* The patient is billed and pays for the procedure. The patient is then reimbursed for expenses by the insurance carrier to the extent designated within the policy. Insurance coverage appears to have begun in 1980 or 1981, though it is difficult to determine exactly when. St. Mary's Hospital, in which Myler does his work, indicates that third-parties have not refused payment, though some have questioned the new procedure. The only difficulty is in being reimbursed for the procedure itself. The hospital visits are reimbursed without question.

A number of major health insurance carriers were contacted. One carrier, New York Life Insurance, said "anything ordered by an M.D. is covered, unless specifically excluded" (21). A second carrier, Mutual of Omaha, stated that their major catastrophic insurance covered "treatment by a physician or surgeon" and "service by a radiologist for diagnosis of treatment" (20).

The fact that PTA is a familiar procedure with proven efficacy makes PTCA more acceptable to carriers. (HCFA has approved payment for the PTA procedure when used in lower extremities.)

Additionally, FDA's decision to release the catheter from the investigational devices list is viewed as important. With various restrictions due to an IDE status, health care facilities are reticent to do PTCAs. During the investigational stage, physicians keep informed of the development of the procedure. Once it is cleared by FDA, acceptance is fairly rapid (see fig. E-2) (20).

Market Factors

Of the approximately 100,000 CABG procedures performed each year, approximately 10 percent of these could be replaced with PTCA. Thus, an estimate of the maximum annual demand of the primary market for PTCA is: 100,000 × 10 percent = 10,000 procedures per year. Additional demand could be realized if indications for the procedure were broadened.

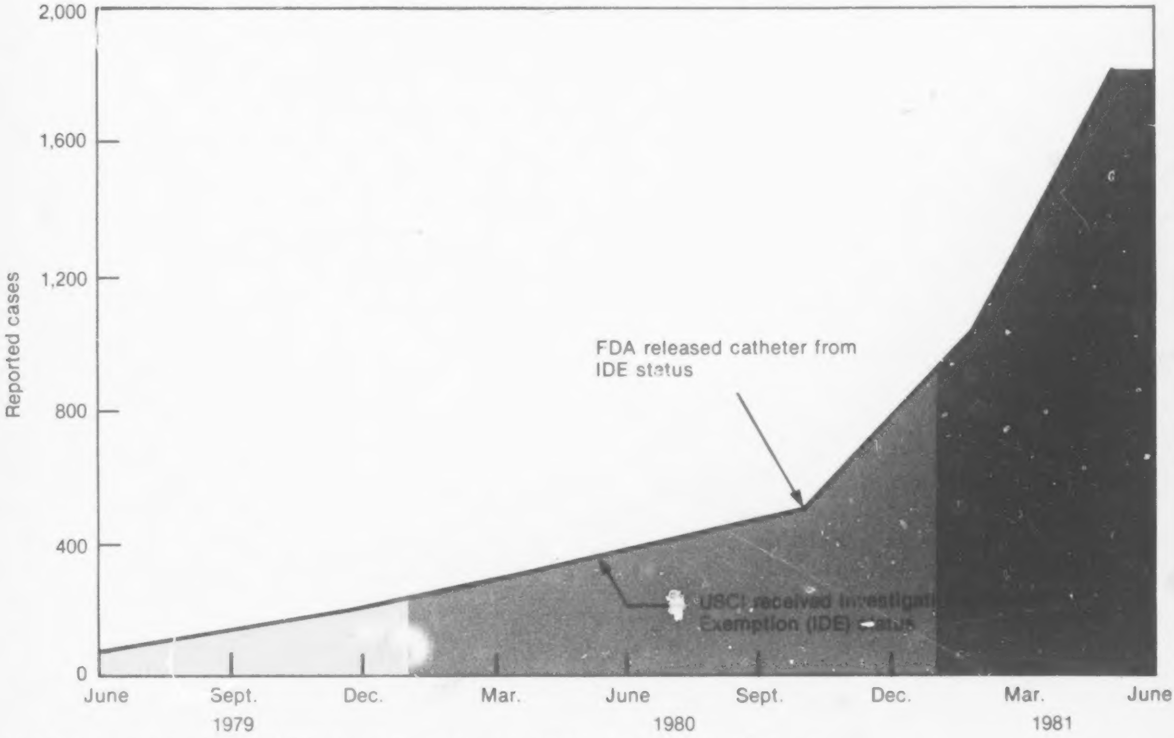
Currently, there are two manufacturers of the special catheter used in PTCA: USCI of Massachusetts and ACS of Santa Clara, Calif. It seems unlikely that additional companies will enter the field. The market is limited in size. Also, the technological barriers to entry are substantial. By 1981, USCI's catheter had

---

*Excluding Blue Cross and Blue Shield.

*Ann, of Dr. Myler's office, personal communication.

Figure E-2.—Reported Cases of PTCA



SOURCE: S. Mullin, Interim Registry of PTCA, Cardiac Disease Branch, National Heart, Lung, and Blood Institute.

been released from IDE status, whereas ACS's catheter had not. ACS expected to receive premarketing approval within 6 months.

The price to the patient of a CABG is quoted as $20,000 at Stanford University Hospital. The price of PTCA is $3,000 (see table E-4). The source of the cost of PTCA was unable to state unequivocally whether the price included a backup team for CABG. However, the patient would be charged $3,000.

Some percent of those patients undergoing PTCA eventually have CABG performed anyway. A true cost comparison should factor this in (see table E-5). The cost savings are such that it would be necessary for 85 percent of the patients who receive PTCA to *also* have CABG before the cost savings advantage of PTCA would be nullified.

Table E-4.—Cost Comparison: PTCA, CABG

CABG	$20,000
PTCA:	
Supplies:	
Catheter	$387
Other catheters	
Medical supplies	500
Other	93
	$980
Personnel: Clerical, technician, two cardiac cath lab nurses (3 hours for procedure)	$74
Benefits	15
	$89
Coronary hemodynamic (partial charge)	$75
Indirect costs, departments and hospital	137
	$1,281
Adjusted for inflation, 15%	1.15
	$1,473
Profit markup, 10%	1.10
	$1,620
Physician fees	1,275
Total cost (price to patient)	$2,895

SOURCE: P. Berry, Cardiac/EKG Department, Stanford University, 1981.

BEST AVAILABLE COPY

Table E-5.—Expected Cost and Break-Even

Expected cost:
Assume: 10 to 15% of PTCAs are followed up with CABGs.
Expected cost per patient:
(3,000) (1.00) + (20,000) (0.10) = $5,000
(3,000) (1.00) + (20,000) (0.15) = $6,000
Expected cost per patient = $5,000 to $6,000.
Break-even:
$3,000 + $20,000 (X) = $20,000
20,000 (X) = 17,000
X = 0.85
where X = percent of patients having PTCA who also need CABG.

Sawi's References

1. Berry, P., Business Manager, Cardiac/EKG Department, Stanford University Hospital, personal communication, June 1, 1981.
2. Brahme, F., et al, "Evaluation of Transluminal Recanalisation of the Femoral Artery," *Acta Chir. Scand.* 135:679, 1969.
3. Colapin, R. F., et al., "Percutaneous Transluminal Angioplasty of Peripheral Vascular Experience: A Two Year Experience," in *Cardiovascular and Interventional Radiology,* E. Zeitler (ed.) (Berlin, Heidelberg: Springer-Verlag, 1980).
4. Dotter, C. T., Oregon State University, Portland, Oreg., personal communication, Apr. 24, 1981.
5. Dotter, C. T., "Transluminal Angioplasty," in *Angigraphy,* vol. 2, 3d ed., H. L. Abrams (ed.) (Boston: Little Brown & Co., 1980).
6. Dotter, C. T., and Judkins, M. P., "Transluminal Treatment of Arteriosclerotic Obstruction," *Circulation,* 30:654, 1964.
7. Finklestein, S., Alfred P. Sloan School of Management, Massachusetts Institute of Technology, personal communication, May 29, 1981.
8. Fogarty, T. J., et al., "A Method for Extraction of Arterial Emboli and Thrombi," *Surg. & Gynec. Obstet.* 116:241, 1963.
9. Grossman, N., Product Manager, Cardiac Catheters, USCI, Billerica, Mass., personal communication, May 29, 1981.
10. Grüntzig, A. R., "Coronary Percutaneous Transluminal Angioplasty: Preliminary Results," *Circulation* 58 (Suppl. 2):56, 1978.
11. Grüntzig, A. R., "Coronary Transluminal Angigraphy," *Circulation* 56 (Suppl. 3):84, 1977.
12. Grüntzig, A. R., "Percutaneous Transluminal Coronary Angioplasty," unpublished document obtained from H. Abrams, Stanford University, April 1981.
13. Grüntzig, A. R., "Percutaneous Transluminal Coronary Angioplasty—Present State of the Art," *Circulation* 60:1031, 1979.
14. Grüntzig, A. R., "Transluminal Dilatation of Coronary Artery Stenosis," letter, *Lancet* 1:263, 1978.
15. Grüntzig, A. R., et al., "Nonoperative Dilatation of Coronary Artery Stenosis: Percutaneous Transluminal Coronary Angioplasty," *N. Eng. J. Med.* 301:61, 1979.
16. Grüntzig, A. R., et al., "Percutaneous Transluminal Coronary Angioplasty: A Cooperative Study," *Am. J. Cardiol.* 43:384, 1979.
17. Kanne, M., Executive Vice-President and Sales Manager, Cook, Inc., Bloomington, Ind., personal communication, May 29, 1981.
18. Katzen, B. T., and Chang, J., "Percutaneous Transluminal Coronary Angioplasty With Grüntzig Balloon Catheter," *Radiology* 130:623, 1979.
19. Mullin, S., Interim Registry of PTCA, Cardiac Disease Branch, National Heart, Lung, and Blood Institute, Bethesda, Md., personal communication, May 14, 1981.
20. Mutual of Omaha, Medical Coverage Representative, personal communication, May 29, 1981.
21. New York Life Insurance, Coverage Representative, personal communication, May 29, 1981.
22. Nordenström, B., "Balloon Catheters for Percutaneous Insertion Into the Vascular System," *Acta Radiologica* 57:411, 1962.
23. Rapaport, E., "Percutaneous Transluminal Coronary Angioplasty—Editorial," *Circulation* 60(5):969, 1979.
24. Schoop, W., et al., "Early and Late Results of PTD in Iliaca Stenosis," in *Percutaneous Vascular Recanalization: Technique, Application, Clinical Results,* E. Zeitler, et al. (eds.) (Berlin, Heidelberg: Springer-Verlag, 1978).
25. Seidlinger, S. I., "Catheter Replacement of the Needle in Percutaneous Angiography," *Acta Radiologica* 39:368, 1953.
26. Simpson, C., Advanced Catheter Systems, Santa Clara, Calif., personal communication, May 22, 1981.
27. Van Andel, G. J., *Percutaneous Transluminal Angioplasty: The Dotter Procedure* (Amsterdam-Oxford: Excerpta Medica, 1976).
28. Wierny, et al., "Long-Term Results in 100 Consecutive Patients Treated by Transluminal Angioplasty" *Radiology* 112:543, 1974.
29. Zeitler, E., "Percutaneous Dilatation and Recanalization of Iliac and Femoral Arteries," in *Cardiovascular and Interventional Radiology,* E. Zeitler (ed.) (Berlin, Heidelberg: Springer-Verlag, 1980).
30. Zeitler, E., et al. (eds.), *Percutaneous Vascular Reconstruction: Technique, Application, Clinical Results* (Berlin, Heidelberg: Springer-Verlag, 1978).
31. Zeitler, E., et al., "Treatment of Occlusive Arterial Disease by Transluminal Catheter Angioplasty," *Radiology* 99:19, 1971.

Maternal Serum Alpha-Fetoprotein

The Federal involvement with MSAFP for detection of fetal neural tube defects is a case study that illustrates successful technology assessment monitoring and interagency coordination The determination of the level of alpha-fetoprotein in maternal serum is the first step in a sequence of diagnostic tests used to screen and diagnose fetal neural tube defects. It aids in detecting two types of defects, anencephalia (absent or undeveloped brain) and open spina bifida (failure of the spine and overlying skin to close over the spinal cord), which together affect 3,000 to 6,000 newborns in the United States each year. Followup procedures, for use when results of screening are abnormal, include repeat serum testing, ultrasonography, and amniocentesis (108).

In late 1978/early 1979, FDA was on the verge of approving a 2-year interim period for widespread usage of MSAFP. However, a special interest group, the spina bifida parents, questioned the quality of the FDA data and the impending diffusion of MSAFP. The Centers for Disease Control (CDC) also became concerned over manufacturers' ability to consistently produce quality components for a marketable kit, as well as over physicians' readiness to work in close cooperation with their patients. Compounded further by ethical and reimbursement issues, CDC sent a formal memorandum to the Office of Health Research, Statistics, and Technology Director, enumerating the concerns with the technology. As a result, the MSAFP screening test was discussed at a meeting of the Technology Coordinating Committee of the Department of Health and Human Services; and NCHCT was asked, through the committee, to coordinate development of departmental policy (108,315).

In November of 1979, former Secretary Harris was briefed, at her request, on departmental activities relating to MSAFP. NCHCT coordinated the briefing. It was agreed that the Public Health Service (PHS), through CDC, should conduct a controlled epidemiologic field study to obtain needed clinical data on MSAFP and to supplement FDA's postmarketing surveillance and data collection program. CDC's protocol was subsequently reviewed by the Technology Coordinating Committee. In July of 1980, NCHCT and FDA cosponsored a national educational conference on MSAFP. By November of 1980, regulatory proposals had been readied for publication (119).

In the November 7, 1980, issue of the Federal Register, FDA published proposed regulations to restrict the sale, distribution, and use of alpha-fetoprotein test kits used in detecting fetal neural tube defects. Also, in that same edition of the Federal Register, CDC and HCFA published jointly proposed regulations pertaining to quality control and proficiency testing for clinical laboratories engaged in alpha-fetoprotein testing. Public hearings on these regulatory proposals were held on January 15-16, 1981. Testimony was presented by some 40 to 50 individuals at these hearings, and approximately 650 written comments were received by FDA and CDC/HCFA subsequent to the publication of the proposed rules. These responses were analyzed and assessed by FDA and CDC/HCFA. FDA found that it had several options in regard to the restricted or unrestricted release for marketing of the alpha-fetoprotein test kits; these were under review as of January 1982 by Arthur Hayes, the FDA Commissioner (120).

Hemodialysis and Kidney Transplant Surgery*

by Katherine R. Jones, Ph.D. (Candidate)
School of Education, Stanford University

Background Information

Hemodialysis and kidney transplant surgery are two alternative forms of therapy for chronic renal failure. Permanent or chronic or end-stage renal disease (ESRD) occurs when an individual irreversibly loses a sufficient amount of kidney function so that life cannot be sustained without treatment intervention. Chronic renal failure may be caused by any of a number of separate diseases (glomerulonephritis, pyelonephritis, polycystic kidney disease, hypertension, diabetes mellitus, and others), but the uremic or ESRD state is their final outcome.

Hemodialysis is a treatment that involves using a machine—the artificial kidney—to achieve the vital functions previously performed by the kidneys. A patient undergoes the treatment from 4 to 6 hours a day, two or three times a week. The treatment can be given in hospital-based dialysis units, freestanding units (for-profit or not-for-profit), or in the patient's home. A fast growing substitute for hemodialysis is continuous ambulatory peritoneal dialysis (CAPD), a technique that uses the peritoneum (lining of the abdominal cavity) to cleanse the blood of its impurities.

The alternative treatment to dialysis is kidney transplantation, which is performed with kidneys from either living related donors or cadaveric donors. The best kidney survival rates are achieved with living related donors, especially siblings.

DEVELOPMENT OF THE TECHNOLOGY

Development of the hemodialysis technology began as early as 1913, when Abel, Rowntree, and Turner performed the first dialysis in animals at the Johns Hopkins Medical School (8,27). This team built their own dialyzing system and coined the expression "artificial kidney." On February 28, 1926, Haas performed the first hemodialysis in a human being, using Hirudin (prepared from leech heads) as the anticoagulant (9). The development of the anticoagulant Heparin greatly enhanced the dialysis procedure, as did the marketing of cellophane for use as the artificial dialyzing membrane. Because of an inability to repeatedly access the bloodstream of people suffering from chronic renal failure, however, the use of hemodialysis was limited to patients with acute, reversible kidney failure.

---

*NOTE: Reference citations for Katherine Jones' case study on hemodialysis and kidney transplant surgery appear on p. 183.

In 1943, Kolff developed the first practical model for human hemodialysis, building an artificial kidney using a rotating drum (27). It took 2 more years before Kolff achieved his first *success* in treating acute renal failure patients with hemodialysis. Independently of Kolff, Alwall in Sweden and Murray in Canada were also developing types of hemodialysis machines, and all three published their experiences at about the same time (9,27).

Use of the artificial kidney for patients with acute renal failure continued until 1960. In that year, a critical technological advance was made when Quinton and Scribner reported the use of a subcutaneous arteriovenous shunt (a plastic tube connected to an artery and a vein in the arm or leg), which allowed repeated access to the circulatory system and thus permitted continuous dialysis treatments (27). After this important development, technological advances were made in the areas of improved blood access devices, dialyzing membranes, dialysis machines, and types of artificial kidneys (8,9).

The first kidney transplant was performed at Harvard by Hufnagal and Hume in 1947 (27). A cadaver kidney was transplanted into the antecubital fossa (area in front of the elbow) of a young woman dying from acute renal failure. The kidney functioned for 2 days, lasting long enough for the patient to regain her own renal function. From 1951 to 1953, Thorn and Merrill referred several patients to Hume, who performed several transplants. They watched patients experience a reversal of their uremic state only to experience a rejection of the kidney in a few days (27).

In February of 1953, the first success was achieved by Hume and his associates. A person survived with a functioning kidney transplant for 5 months and 25 days and demonstrated the potential of the procedure (21). This case also marked the end of experimental transplantation without the use of immunosuppression to prevent graft rejection (21,27).

In 1954, the first transplant between monozygotic twins was performed by Murray and his associates in Boston (2). This case demonstrated that monozygotic twins were, indeed, immunologically identical. Within 5 years, this group had performed eight transplants between twins and had perfected the surgical technique of retroperitoneal iliac fossa (groin area) placement of the graft (2). The initial transplant recipient lived for 8 years before dying of myocardial infarction (21,27).

Clinical and animal experiments with different forms of immunosuppression occurred in 1958 and 1959. Total body irradiation usually proved to be fatal and was soon abandoned (2). Schwartz next introduced drug-induced immunological tolerance, first using mercaptopurine, then switching to a superior derivative called Imuran, a drug still in use today (14,27). Steroids joined Imuran in treatment of graft rejection in 1962 (2). In 1963, Terasaki began using serotyping techniques to select immunologically favorable kidney donors (2). New tissue-matching techniques, organ acquisition and preservation procedures, and immunosuppressive drugs have since been introduced, although improvement is still needed in long-term organ survival rates (21,27).

DIFFUSION OF THE TECHNOLOGY

In the early 1960's, hemodialysis and kidney transplant became accepted as life-extending therapies for victims of chronic renal failure. In America, much of the research was initially supported by the John A. Hartford Foundation and later by the Artificial Kidney/Chronic Uremia Program (AKCUP) of the National Institutes of Health (NIH). AKCUP in the National Institute of Arthritis, Metabolism, and Digestive Diseases (NIAMDD) was founded in 1965 with a contract research program to build a better artificial kidney. This program was mandated by the House and Senate Appropriations Committees 1 year *after* the artificial heart program, although the artificial kidney was more developed at the time (27).

The development of hemodialysis as a mode of therapy posed complications for NIH. NIH found it difficult to support Scribner's development of clinical applications of the artificial kidney and was never prepared to do so on the scale he requested (2). The NIH orientation toward biological and biochemical processes led to a preference to fund research on kidney disease etiology, not the clinical applications of that research. Funding for transplant research was not as problematic, since those involved with immunological research were well known as basic researchers (27).

In 1964, the Senate Appropriations Committee stated that PHS had the authority to provide demonstration and training funds for artificial kidney programs. In 1965, PHS established the Kidney Disease Control Program (KDCP), which funded 14 community treatment centers around the country and demonstrated the organizational feasibility of dialysis in various settings. Although these contracts were gradually phased out beginning in 1968, many PHS-funded dialysis centers became nationally prominent hemodialysis provider institutions (27).

In 1969, KDCP became part of the Regional Medical Program (RMP), and the emphasis shifted from demonstrations of feasibility to the building of dialysis *capacity* (27). This was accomplished through centralized funding and policy control in Washington, and decentralized funding of facilities through the RMP agencies (27).

A very significant role in the diffusion of dialysis technology was played by the Veterans Administration (VA). In 1963, 2 years before the initiation of AKCUP and KDCP, VA announced its intention to establish dialysis centers in 30 VA hospitals (27). It proceeded to do so over the next several years. Dialysis and transplantation were provided for all qualified veterans with chronic kidney failure, whether or not service-connected. Public Law 89-785 (Nov. 7, 1966) provided that nonveterans could also receive such services from VA hospitals if facilities were available, but that VA had to be reimbursed the full cost of services rendered under agreements with other hospitals (27,31).

By 1971, VA had initiated a home dialysis program, had opened its first home training unit, and had also initiated satellite dialysis. By 1972, VA was dialyzing 25 percent (979 patients) of the Nation's dialysis patients in 44 treatment centers and another 26 patients in branch centers. As of March 1973, the VA system had 501 dialysis beds (including 123 for home training) and 10 hospitals that were operating branch dialysis centers. VA had also provided backup medical services for 766 patients being dialyzed at home. Thirty-three VA hospitals reported a total of 327 transplants performed in 1972.

The activity in PHS and VA related to hemodialysis and transplantation prompted the Government to conduct a high-level policy review of the situation. The Bureau of the Budget established the Gottschalk committee in 1965 to review the implications of therapy for ESRD for the entire Nation (27). In 1967, the Gottschalk committee (12) declared that hemodialysis and transplantation were acceptable forms of therapy and recommended that a national treatment benefit program be established by amending title XVIII of the Social Security Act. No legislation was enacted as a result of this report, partly because of the simultaneous release of a PHS report that emphasized research and prevention rather than treatment.

Gottschalk (12) estimated that in 1962, one out of every five patients dying from chronic uremia was medically suitable for dialysis and transplantation. Of the 7,000 *new* renal patients in 1968 who *would* be suitable for treatment, transplants were available for approximately 450 and chronic dialysis for approximately 550 (12). The treatment technology was available but was prohibitively expensive. In 1965, the costs for dialysis and transplant were as follows (12):

- *home dialysis:* average $6,000 per year, range $3,750 to $9,800;
- *in-center/hospital dialysis:* average $10,000 per year, range $8,400 to $21,000; and
- *transplant:* average $13,300 surgery and recovery plus $200 to $1,000 per year, range $10,000 to $22,000.

The eight primary sources of funding for dialysis and transplantation prior to the establishment of the ESRD program were the following (7):

1. *Medicare.*—This program helped finance treatments for persons 65 and over (but few patients this age were selected for treatment).
2. *Medicaid.*—Under this program administered by the Social and Rehabilitation Service, the States could assist in paying for kidney disease treatment. Eligible individuals were those who received public assistance under certain titles of the Social Security Act, or those persons in certain States whose income and resources were insufficient to meet medical needs. Services varied among the States.
3. *Vocational Rehabilitation Act.*—Using Federal grants to the States, this program helped 300 to 400 individuals with kidney disease annually at a cost of about $1 million.
4. *Comprehensive Health Planning (authorization expired June 30, 1974).*—According to the Department of Health, Education, and Welfare, RMP of the Health Resources Administration was, as of April 1975, investing $4.8 million in grants to develop chronic kidney disease treatment services reimbursable under the Medicare program. Funds were primarily aimed at startup costs of cadaver kidney procurement systems and specialized laboratory services.
5. *Military.*—The Army, Navy, and Air Force provided a limited number of dialysis machines for home and center programs through the Civilian Health and Medical Program of the Uniformed Services.
6. *Research.*—NIAMDD was the NIH component primarily responsible for supporting kidney disease research. The National Institute of Allergy and Infectious Diseases (NIAID) of NIH supported research on the infectious and immunological aspects of kidney disease and was the major source of funds for transplantation.
7. *State and Private Involvement.*—A General Accounting Office (GAO) survey of 14 States in 1973 found 8 States that were appropriating funds to directly assist patients with kidney disease. Wide variation existed between the States: One

State legislature appropriated about $4,400 per 100,000 population, while another appropriated about $11,000 per 100,000 population. Two of the States operated their own dialysis facilities.

8. *Private Sources.*—Private sources included out-of-pocket payments, savings withdrawals, private health insurance, the National Kidney Foundation, local community organizations and fund drives, and employer contributions.

ESRD Program

From the mid-1960's until 1972, a long policy debate occurred over who should be responsible for funding treatment of patients with chronic renal failure. The issue was publicized in the media, and the existence of "death committees" (groups of physicians or health professionals who decided who would be dialyzed and who would be allowed to die) became well known (18). There was mounting pressure for the Federal Government to take action that would relieve the patients and other payers of the expensive burden of this lifesaving technology.

The debate culminated in 1972 with the passage of section 2991 of Public Law 92-603 of the Social Security Amendments of 1972, a law that extended coverage for renal disease treatment to over 90 percent of the population (30). Medicare eligibility began in the third month after the month in which a course of dialysis was initiated and ended in the 12th month after the month in which a person had a functioning kidney transplant. Factors that led to the congressional decision to pay for ESRD treatment included a recognition that the alternative to life sustainment by dialysis was death, that ESRD treatment was very expensive, and that there occurred 7,000 to 10,000 uremic deaths a year because of the limited availability of dialysis facilities.

HCFA assumed responsibility for the ESRD program in 1978. HCFA prescribes standards for treatment by its regulations and approves payments for services through local insurance companies or other intermediary agencies. HCFA is also responsible for the quality of care that patients receive and exercises that responsibility through existing National, regional, and State agencies.

CHANGES IN THE PATIENT POPULATION

There have been changes in the number of patients receiving transplants and the proportion of patients receiving home dialysis since institution of the ESRD program.

Transplants.—Between 1963 and 1972, the number of kidney transplants had been increasing steadily, going from 163 to 1,993 a year (7). After 1972, the number grew at a slower pace, and it plateaued in 1975 (10,26) (see table E-6).

The reasons for the decline in transplants included lack of improvement in graft success rates (although patient survival rates *had* improved), decreased donor pool due to smaller families, and financial disincentives in the Medicare regulations. Benefits ended the 12th month after transplant surgery. If a person lost his or her kidney after this time period, that person had to undergo another 3-month waiting period before Medicare began paying for the resumed dialysis treatments. The patient also had to pay the costs associated with transplant failure. In 1975, GAO (7) recommended that the waiting period and associated disincentives for transplant, a less costly treatment modality, be eliminated from the law. These recommendations were implemented in 1978.

Hemodialysis.—The number of patients receiving hemodialysis grew rapidly after 1970. From 11,000 dialysis patients in 1973, the program expanded to about 50,000 dialysis patients in 1980 (see table E-7). This growth occurred primarily as a result of changes in patient selection criteria. Originally, selection was limited to patients 15 to 45 years of age who were in good health apart from their renal disease. Today, there is essentially one criterion for acceptance—

Table E-6.—Number of Kidney Transplants Performed in the United States, 1951-79

Year	Number[a]	Percentage change
1951-62	75	
1963	163	117.3%
1964	239	46.6
1965	305	27.6
1966	338	10.8
1967	448	32.5
1968	676	50.9
1969	838	24.0
1970	1,091 (1,460)	30.2
1971	1,616 (2,909)	48.1
1972	1,993 (2,852)	23.3
1973 (Medicare coverage)	3,017	51.4
1974	3,190	5.7
1975	3,730	16.9
1976	3,504	−6.1
1977	3,973	13.4
1978	3,949	−0.6
1979	4,271	8.2

[a] Numbers in parentheses reflect discrepancies in the literature.

SOURCES: Comptroller General of the United States, *Treatment of Chronic Kidney Failures: Dialysis, Transplant Costs, and the Need for More Vigorous Efforts,* 1975; Office of Special Programs, Office of End-Stage Renal Disease, Health Care Financing Administration, *End-Stage Renal Disease Second Annual Report to Congress, Fiscal Year 1980,* 1980; and E. A. Freedman, et al., "Pragmatic Realities in Uremia Therapy," *N. Eng [illegible] Med.* 298:368, 1978.

Table E-7.—ERSD Patient Population, 1972-80

Year	Number of hemodialysis patients	Percentage of patients on home dialysis[a]	Number of patients on CAPD[b]	Percentage of patients over 65 years	Number of dialysis facilities
1972	10,000	40%	—	NA	NA
1973	11,000	35.9%	—	NA	NA
1974	18,875	32.7%	—	5%	664
1975	22,000	28%	—	11%	NA
1976	30,131	23.7% (13%)	—	17%	700
1977	32,435	11.56% (20%)	—	20%	761
1978	36,463	12.42%	8	19%	850
1979	45,565	13% (10%)	1,800	NA	950
1980	50,000	NA	3,000	NA	1,000

[a] Numbers in parenthesis reflect conflicting reports in the literature.
[b] Medicare coverage began in 1978.
NA = not available.

SOURCES: Office of Special Programs, Office of End-Stage Renal Disease, Health Care Financing Administration, *End-Stage Renal Disease Second Annual Report to Congress, Fiscal Year 1980*, 1980; R. Rettig, *Implementing the End-Stage Renal Disease Program of Medicare*, 1980; and Health Services Administration, *End-Stage Renal Disease Medical Information System—Facility Report Number 1*, 1976.

disabling uremia. A large proportion of the dialysis population is either very young or very old and suffers from other serious diseases, such as liver disease, cancer, and diabetes. Table E-7 also shows that there has been a fairly rapid *decrease* in the proportion of home dialysis patients since the beginning of the ESRD program.

When the Social Security Amendments of 1972 were passed, 40 patients per million were receiving long-term hemodialysis treatment in the United States, almost entirely under the auspices of nonprofit organizations (24). The number of patients now receiving dialysis treatments in the United States exceeds *200* per million population, an eightfold increase, and is the highest in the world (24).

The composition of patients receiving dialysis treatment in the United States has also changed since 1972. The average age of the maintenance dialysis population has increased. The mean age of dialysis patients rose from 42 in 1970 to 50 in 1977 (18,31). Americans well past retirement age may be placed on dialysis. A survey of European dialysis centers in 1977 revealed that 70 percent had no age limit to dialysis; 22 percent as a general rule excluded patients over 65; and 8 percent excluded patients over age 55 (1).

CHANGES IN THE USE OF THE TECHNOLOGY

The changes in the treated renal patient population have occurred mainly as a result of increased availability of funding for dialysis and transplantation. The increasing number of new patients presenting each year, increasing average age of new patients, and increasing number of patients with serious complications have contributed to the decreasing use of home dialysis and transplantation and to an overall rise in morbidity and mortality.

There has been much national concern expressed over the declining proportion of home dialysis patients, since significantly lower costs are associated with home treatment, especially after the first year. The Medicare regulations themselves included disincentives for home dialysis (3,7). For example, the Medicare regulations did *not* require centers to provide training programs for home dialysis (7). Instead, they required more out-of-pocket costs for home dialysis supplies and equipment and did not provide reimbursement for the services of a home dialysis assistant nor for the effort involved in renting equipment, ordering supplies, and other bookkeeping requirements. Home patients incurred additional costs for home modification and higher electric and water bills (13,19).

Some of the movement back to facility dialysis was the result of the stresses on family life caused by home dialysis. Opponents of home dialysis have stated that many patients were initially placed on home dialysis solely because of limited funds for facility dialysis treatment and that Medicare has removed these financial barriers to the preferred treatment setting. Other factors in the home dialysis/facility dialysis controversy include the personal philosophy of the physician or hospital treating the patient, increased age and morbidity of dialysis patients that reduce their suitability for home treatment, and the impact of facility proprietary-status on the location decision outcome.

Proponents of home dialysis have argued that the costs of home dialysis are lower than those for facility dialysis, that quality of life is improved as patients are more independent and able to work, and that morbidity and mortality rates are lower (5). Using data from 1972, the NIH National Dialysis Registry reported a home dialysis 3-year mortality rate of 21.4 percent and a facility dialysis 3-year mortality rate of 28.6 percent (13). Using 1976 cumulative data, it reported the following annual death rates: home dialysis, 6.7 percent; freestanding-unit dialysis, 7.5 percent; hospital-based-unit dialysis, 10 percent.

GAO in 1977 reported that mortality rates for dialysis patients were unavailable to it and that GAO was therefore unable to compare mortality rates of home-treated v. center-treated dialysis patients (6). Neither the Medicare billing system nor the ESRD medical information system recorded morbidity data (4). The ESRD amendments of 1978 mandated HCFA to collect such data, and information on hospital admissions, average length of stay by disease category, survival rates, and program costs associated with alternative treatments of ESRD was to be reported in 1981.

GAO *was* able to collect cost information and in 1975 reported the following data based on 1972-73 cost or charge information (7):

Type of dialysis	*Average yearly cost*	*Range*	*Per treatment cost*
Home	$14,900—first year	$9,300-$22,200	$96
	$7,000—following years	$3,900-$10,300	$46
Freestanding	$27,600	$16,140-$41,000	$203
Hospital-based	$30,500	$11,500-$49,100	$202

NIH cost estimates (excluding physician fees) based on 1973 cost data (5 dialysis centers) were as follows (10):

Type of dialysis	*Average yearly cost*	*Per treatment cost*
Home	$15,000—first year $6,500—following years	$33-$66
Freestanding	$16,520	$100-$116
Hospital-based	$24,736	$146-$259

In 1973, the Medicare reimbursement rate was $150 for center dialysis and $50 for home dialysis, both including physicians' fees and assuming 156 treatments a year.

In 1977, GAO estimated the following costs for home dialysis (6):

First year	
Equipment	$6,800
Training—24 treatments @$158	3,792
Physician fees (training)	500
Physician fees (supervision)—12 months @$140	1,680
Backup dialysis—16 treatments @$138	2,208
Supplies and equipment—116 treatments @$55	6,380
Reasonable charges covered by Medicare	$21,360
Second year	
Backup dialysis—19 treatments @$138	$2,622
Physician fees—12 months @$140	1,680
Supplies and equipment—137 treatments @$55	7,535
Reasonable charges covered by Medicare	$11,837

A study by Roberts, et al. (29), estimated the following costs for dialysis and transplantation in 1980:

Home dialysis—first year	$22,760
Home dialysis—remaining years (per year)	13,237
In-center dialysis (per year)	24,800
Cadaver transplant—first year	23,400
Cadaver transplant—second year	3,000
Cadaver transplant—third year	1,500
Cadaver transplant—remaining years (per year)	750
Related donor transplant—first year	20,700
Related donor transplant—second year	1,500
Related donor transplant—remaining years (per year)	500
Graft rejection (per year)	9,000

For the years 1976 through 1979, the ESRD program (23) estimated the following kidney acquisition costs (donor surgery costs):

Year	*Average*	*Low*	*High*
1976	$4,223	$1,003	$13,197
1977	4,690	1,000	15,000
1978	5,790	1,000	12,683
1979	5,906	1,000	15,000

Average transplant charges in 1973 were $12,800, with a range of $5,500 to $20,500. Average transplant charges in 1979 were $23,000, and cadaveric kidney transplants averaged $25,000 in 1978 (30).

The above cost per year figures show that substantial savings could be achieved by shifting more patients to treatment by home dialysis or transplant (30). Roberts and his associates reported in 1980 that living, related donor transplants were the least costly treatment modality and had the greatest survival time; center dialysis (hospital-based) was the least cost effective (29). Shifts to either home dialysis or cadaveric donor transplantation would save from $7,000 to $8,000 per life year or $284 million per year for the existing ESRD program (29).

GROWTH OF PROPRIETARY FACILITIES

More than 20 percent of the Nation's hemodialysis patients are now dialyzed in for-profit units. Proprietary facilities present difficult ethical problems for the medical community (20). Physicians who render care to a patient also share in the profitability of that function. For-profit dialysis units tend to prefer maintaining patients in an outpatient setting rather than a home setting. There are many cost disincentives to home hemodialysis. Since home dialysis means less profit, profit may influence the choice of mode of care (20). On the other hand, proprietary facilities filled a need when university hospitals and the Government could not or would not expand facilities. Such facilities are extremely cost effective, and all function within the Medicare reimbursement screen. Freestanding units can deal with Medicare billing, private insurers, and other funding sources more efficiently than hospital billing departments (3). They also have financial incentives to maintain a high patient census and to provide short, highly efficient dialysis using ultrafiltration technology in order to maximize the business efficiency of the facility. Physician reimbursement for services also gives for-profit units an incentive to encourage facility dialysis, since home dialysis generates fees at the office visit level only (3).

Function and Structure of the ESRD Program

REIMBURSEMENT FOR PHYSICIANS' SERVICES

Hemodialysis.—Under the ESRD program, there are two methods of reimbursement for physicians' services: initial method and alternative payment method. By the *initial method,* physicians are paid directly by the facility for their supervisory services during dialysis. For other nonroutine services required by the patient, physicians bill on a fee-for-service basis. The average payment rate to the physician for supervisory services during dialysis is part of the overall dialysis charge and averages $13 per treatment ($12 for free-standing units). By the *alternative reimbursement method,* physicians are paid a monthly fee for each patient for the full renal care of that patient. Such care includes supervisory services during dialysis plus all other related services furnished during a particular month. Prior to July 1, 1978, alternative monthly allowances for physician services to patients dialyzing in facilities ranged from a minimum level of $160 to a maximum level of $240. Allowances for physician services for treatment of patients dialyzing at home ranged from a minimum of $112 to a maximum of $168. These amounts are subject to Medicare Part B coinsurance and remained constant from time of their implementation in 1974 to July 1, 1978.

The monthly allowances were increased on July 1, 1978, to reflect changes in the customary and prevailing charges for internists and routine followup office visits, but were not to exceed the increase in the medical care index, which rose 20.9 percent from July 1975 to July 1978. The resulting revised monthly payments to the 957 physicians using the alternative method of payment now range from $180 to $260 before coinsurance for facility patients and from $126 to $182 for home patients. The price index adjustment resulted in an arithmetic mean payment before coinsurance of $220 a month for facility patients and $154 a month for home patients.

Transplant.—All physicians' fees are covered as follows:

- 1978-79—Excise surgeon (donor kidney): $350 for 1 kidney, $700 for two kidneys.
- 1979—Transplant surgeon: $1,600 to $2,500, depending on number of services provided.
- 1979—Transplant surgeon: $1,690 to $2,730, depending on number of services provided.

ESRD NETWORKS

ESRD networks in 32 geographic areas covering the United States were established by regulations published on June 30, 1976. Their role and function were restated in section 1881 of Public Law 95-292. In 1977, Federal funding policies were changed to encourage network organizational efforts. The year 1978 was devoted to the establishment of viable organizations and the conduct of multiple functions: inviting Medicare-approved ESRD facilities to join the network coordinating councils, developing operating rules and procedures, and hiring professional and technical staff personnel. All 32 networks had secured the services of an executive director by September 1978.

Medical review boards began in 1978 to fulfill their regulatory functions (14): 1) monitoring the effect of long-term programs by assessing appropriateness of patients for proposed treatment procedures, 2) reviewing the comparative performance of facilities and physicians for areas of patient care, 3) conducting medical care evaluation studies, and 4) performing other studies as needed.

Network coordinating councils review and recommend approval/disapproval of applications for new or expanded dialysis facilities, a process that has led to charges of conflict of interest. HCFA regional offices consider the network recommendations together with recommendations from the health systems agency and State health planning agencies in making their facility certification decision. In many networks, 75 to 80 percent of the facility review committee may have a direct proprietary interest in the decision being made.

The network coordinating councils were requested by HCFA in November/December 1978 to develop goals relating to self-dialysis and transplantation. In mid-January, they were asked to submit interim statements of 1979 goals to HCFA. As of March 15, 1977, 12 of the 32 networks had submitted statements. According to the 1980 annual report of the ESRD program, 24 of the networks had established goals for self-dialysis training, home dialysis, or for self-dialysis programs and kidney transplantation (23). Eight networks refused or evaded the HCFA operating guidelines: Network 3 (Northern California), 6 (Arizona and New Mexico), 15 (Illinois), 23 (Washington, D.C.), 24 (Delaware), 27 (Connecticut), 28 (Maine, Vermont, New Hampshire, Massachusetts, and Rhode Island), and 32 (New Jersey).

DIALYSIS PAYMENT RATES (22,23)

In 1979, the average payment rate was $149 per treatment. This is a combined weighted average of payments to hospital-based and non-hospital-based units. Hospital facilities were paid the lesser of their costs or a national payment limit (the screen), and the average payment was $159 per treatment. Independent facilities were paid the lesser of their charges or the national payment limit, and the average payment was $138 per treatment.

Any ESRD facility desiring a payment rate above the national limit had to request a reimbursement exception and submit documentation of its higher costs. In 1978, HCFA approved 208 ESRD facility requests for reimbursement exceptions. Some of the primary reasons for approval included:

- treatment of an unusually ill population;
- treatment of an unusual patient population (e.g., children);
- location in a high-cost or low-utilization area;
- recent approval and continuous experience of low utilization;
- demonstrated low utilization due to sporadic workload (referral hospital).

The number and range of payments approved in excess of program payment screens in 1978 and 1979 were as follows:

Range	*1978*	*1979*	*Home training, 1979*	*Peritoneal*
Up to $150	28	27	0	0
$151-$170	63	80	8	0
$171-$190	69	62	15	1
$191-$210	30	35	20	4
$211 +	208	221	52	5

According to the 1980 ESRD annual report, however, HCFA actually approved 278 reimbursement exception requests (in full or in part) in 1979, while denying or returning another 23 requests (23).

GROWTH IN ESRD PROGRAM COSTS

The cost of the ESRD program grew from $250 million in 1974 to over $1 billion in 1979, greatly exceeding original congressional estimates of potential costs (29). However, the number of patients receiving treatment also exceeded estimates. According to Kolata (18), when inflation is taken into account, the costs of dialysis *per patient* have decreased since the program began. (But total published costs underestimate true costs to the Government because a substantial number of patients collect Federal disability payments (16).)

In March 1977, the average weekly payment for the ESRD program was $6 million; in July 1978, it was $10 million a week; and by August 1978, it had reached $12 million a week (20). Between 1973 and 1977, the annual cost per patient had increased at less than half the annual rate of inflation. From 1977 to 1978, the per capita payment increased from $15,295 to $16,300, a 6.5-percent increase.

However, growing concern has been expressed over the large costs for a program benefiting a rather small number of people (29). In 1979, benefit payments for ESRD exceeded 5 percent of total Medicare expenditures, and were fully 10 percent of expenditures from the Supplemental Medical Insurance fund (Part B) of Medicare, although renal patients comprise only 0.2 percent of the Medicare population (4,29). In addition, fully one-third of ESRD beneficiaries were eligible for Social Security monthly disability benefits. More specifically, 10 percent of Part B funds went to 50,000 ESRD beneficiaries, while 90 percent of the funds went to 23 million elderly enrollees; as of 1977, the ESRD patients received $13,555 per capita, while the elderly enrollees received $218.37 per capita (28). The Government set a fee of about $28,000 per year for each patient in an outpatient facility in 1979. Medicare paid 80 percent and the rest was covered by the States and private insurance carriers or was absorbed by the centers. Table E-8 shows the growth in program costs since the beginning of the ESRD program.

Quality of Life on Dialysis

Now that physicians no longer have to make painful choices regarding who gets selected for dialysis, the question has been raised whether too many people in the United States are now being dialyzed. Blagg and Scribner (4) believe that for an ever-increasing proportion of dialysis patients, the quality of life is unacceptable and increasingly costly. Many patients now being accepted into dialysis programs have such severe complicating illnesses that they may be unable to live at home. Known examples include a blind diabetic with severe angina and a nursing home patient who gets transported by ambulance twice weekly to the dialysis center (18). Quality-of-life issues also relate to the "sick room" atmosphere of hospital-based dialysis units (9). Freeman (11) recommends removing dialysis units from hospitals to better designed, less expensive, more cheerful, freestanding units, with the hope that the "mass production line" approach to dialysis can be avoided.

There are now 55,000 patients on hemodialysis. An informal survey of 21 dialysis centers found that 44

Table E-8.—Growth in ESRD Costs, 1974-80

Year	Per capita costs	Program costs (millions)[a]
1974	NA	$250 (283)
1975	$11,600	330 (450)
1976	12,300	450 (598)
1977	16,800	600 (722)
1978	17,300	737 (947) (1.1 billion)
1973	23,500	850.5 (1.2 billion)
1980	28,000	NA

[a] Numbers in parentheses reflect discrepancies in the literature.

NA = Indicates not available.

SOURCE: R. Rettig, *Implementing the End-Stage Renal Disease Program of Medicare*, 1980.

percent of the dialysis patients were not working and that more than 50 percent of this population were probably too sick to work. Twenty percent of the nondiabetic patients were unable to care for themselves completely, and 50 percent of the diabetic patients were unable to care for themselves completely. Skewed incentives may have encouraged insufficient discrimination in the selection of candidates for dialysis (15).

In Britain, many physicians decide not to refer certain types of patients for hemodialysis because of their belief that it is inappropriate treatment for the situation. "In the absence of personal financial incentives to treat more patients with dialysis, the NHS [National Health Service] doctor is more free to decide that these extraordinary procedures for prolonging life do not confer a good enough quality of life to make them suitable for all patients dying of renal failure. Not to treat may be kinder and wiser." (1).

Generalizations About the ESRD Program

According to Rettig (28), most of the problems of the ESRD program have arisen from its administrative system, the planning and operational stages of its implementation, and the substance of reimbursement and medical issues. The most important reimbursement policy has been the screen, or de facto ceiling, on the per treatment reimbursement of outpatient maintenance dialysis. This screen has provided a strong incentive to cost containment. On the other hand, the financial disincentives to home dialysis have been one of the last defensible aspects of reimbursement (28). Rennie (26) has stated that all the problems with the ESRD program can be attributed to inappropriate economic incentives and inappropriate economic deterrents in the present law and regulation.

Blagg and Scribner (3) have identified several problems with respect to the ESRD program's implementation: lack of effective leadership, no continuity in administrative policy, slow and haphazard implementation; increased vulnerability to political lobbying, piecemeal regulations, and lack of meaningful data. The absence of data has made it impossible to assess and compare the quality of care and patient outcomes. It has been impossible to pull out of HCFA's computers data such as percentages of patients dialyzed at home and at centers and such patients' relative mortality rates, ages, and illness levels.

Three important innovations have been introduced by the ESRD program (28). One is the screen on facility reimbursement, which creates strong incentives for delivering outpatient dialysis. The second is the process of reimbursing physicians indirectly by a monthly capitation method, which departs from the traditional fee-for-service reimbursement. The third innovation is the facility certification process, which permits the number and capacity of treatment facilities to increase in reasonable relation to the growth in the patient population.

Jones' References

1. Anonymous, "Selection of Patients for Dialysis and Transplantation," *Br. Med. J.* 6150:1449, 1978.
2. Bergan, J. H., "History of Renal Transplant Registry," in NIAID, *U.S. Kidney Transplant Fact Book* (Washington, D.C.: Department of Health, Education, and Welfare, 1972).
3. Blagg, C. R., and Scribner, B., "Long-Term Dialysis: Current Problems and Future Prospects," *Amer. J. Med.* 68:633, 1980.
4. Blagg, C. R., and Scribner, B., "Medicare End-Stage Renal Disease Program: More Than A Billion Dollar Question," *Ann. Inter. Med.* 93:501, 1980.
5. Cestero, R., "A Regional End-Stage Renal Disease Program: Twelve Years' Experience," *Ann. Inter. Med.* 93:494, 1980.
6. Comptroller General of the United States, *Request To Update Information in the 1975 GAO Report* (Washington, D.C.: General Accounting Office, 1977).
7. Comptroller General of the United States, *Treatment of Chronic Kidney Failure: Dialysis, Transplant Costs, and the Need for More Vigorous Efforts* (Washington, D.C.: General Accounting Office, 1975).
8. Czaczkes, J. W., and De-Nour, A. K., *Chronic Hemodialysis as a Way of Life* (New York: Brunner/Mazel, 1978).
9. Drukker, W., et al. (eds.), *Replacement of Renal Function by Dialysis* (The Hague: Mortinus Nijhaff Medical Division, 1978).
10. Freedman, E. A., et al. "Pragmatic Realities in Uremia Therapy," *N. Eng. J. Med.* 298:368, 1978.
11. Freeman, R., "The Social Impact of Chronic Maintenance Hemodialysis," in *Replacement of Renal Function by Dialysis*, W. Drukker, et al. (eds.) (The Hague: Mortinus Nijhaff Medical Division, 1978).
12. Gottschalk, C. W., *Report of the Committee on Chronic Kidney Disease* (Washington, D.C.: Bureau of the Budget, 1967).
13. Greenberg, D., "Renal Politics," *N. Eng. J. Med.* 298:1427, 1978.
14. Gutch, C. K., and Stoner, M., *Review of Hemodialysis for Nurses and Technicians* (New York: C. V. Mosby, 1979).
15. Gutman, R., et al. "Physical Activity and Employment Status of Patients on Maintenance Dialysis," *N. Eng. J. Med.* 304:309, 1981.
16. Hampers, C., and Hager, E., "The Delivery of Dialysis Services on a Nationwide Basis," *Dialysis & Transplant.* 8:417, 1979.
17. Health Services Administration, *End-Stage Renal Disease Medical Information System, Facility Report Number 1* (Washington, D.C.: Department of Health, Education, and Welfare, 1976).
18. Kolata, G. B., "Dialysis After Nearly a Decade," *Science* 208:473, 1980.

19. Kolata, G. B., "NMC Thrives Selling Dialysis," *Science* 208:379. 1980.
20. Moore, F. D., "End-Stage Renal Disease and the Government," *Dialysis & Transplant.* 7:258, 1978.
21. Moore, F. D., *Transplant: The Give and Take of Tissue Transplantation* (New York: Simon & Schuster, 1972).
22. Office of End-Stage Renal Disease, Health Care Financing Administration, *End-Stage Renal Disease Program Annual Report to Congress, 1979* (Washington, D.C.: Department of Health, Education, and Welfare, 1979).
23. Office of End-Stage Renal Disease, Health Care Financing Administration, *End-Stage Renal Disease Program Second Annual Report to Congress, 1980* (Washington, D.C.: Department of Health and Human Services, 1980).
24. Relman, A., "The New Medical-Industrial Complex," *N. Eng. J. Med.* 303:963, 1980.
25. Rennie, D., "Home Dialysis and the Costs of Uremia," *N. Eng. J. Med.* 298:393,1978.
26. Rennie, D., "Renal Rehabilitation—Where Are the Data?" *N. Eng. J. Med.* 304:351, 1981.
27. Rettig, R., "End-Stage Renal Disease and the "Cost" of Medical Technology," in *Medical Technology: The Culprit Behind Health Care Costs?* S. Altman and R. Blendon (eds.) (Washington, D.C.: Department of Health, Education, and Welfare, 1977).
28. Rettig, R., *Implementing the End-Stage Renal Disease Program of Medicare* (Santa Monica, Calif.: Rand Corp., 1980).
29. Roberts, S., et al., "Cost-Effective Care of End-Stage Renal Disease," *N. Eng. J. Med.* 92:243, 1980.
30. Stange, P., and Summer, A., "Predicting Costs and Life Expectancy for End-Stage Renal Disease," *N. Eng. J. Med.* 298:372, 1978.
31. Steinman, T. I., letter to the editor, *N. Eng. J. Med.* 298:1086, 1978.

Appendix F.—Model for an Institute for Health Care Evaluation*

by John P. Bunker, M.D., and Jinnet Fowles, Ph.D.
Division of Health Services Research,
Family, Community and Preventive Medicine,
Standard University School of Medicine

Rationale

Current Federal policy is to reduce the Government's responsibility for health care, substituting wherever possible market mechanisms, and to vest residual control in regional and local authorities. Towards this end, the Reagan administration has recommended to Congress sharp reductions in expenditures for medical technology assessment. This approach is reflected in the failure to fund the National Center for Health Care Technology (NCHCT) and major cutbacks for the National Center of Health Services Research (NCHSR), the National Center for Health Statistics, and the Office of Research and Demonstrations of the Health Care Financing Administration (HCFA).

The budget of the National Institutes of Health (NIH) has been relatively spared, but, with even modest decreases in NIH funding, any cutbacks can be expected to occur primarily in the areas of evaluation and clinical trials (4). This reduction would occur at a time when the pressures for more comprehensive evaluation are increasing, both from academic institutions and from private and governmental insurers. One partial solution to this conflict might be to develop a private Institute for Health Care Evaluation (IHCE), which would operate as a nonprofit corporation (perhaps replacing NCHCT) and extend the Nation's capacity to evaluate medical technologies.

IHCE could be composed of members from several groups concerned with the evaluation of health care: governmental insurers (HCFA); private medical insurers (Blue Cross, Blue Shield, and commercial carriers), health maintenance organizations (HMOs); professional associations (represented, perhaps, by the Council of Medical Specialty Societies and its program for clinical procedure review); and health consumers. Each of the parties could benefit from the data that IHCE generated. Health care professionals could use the data to improve the quality of patient care; health consumers could have increased information on which to base their selection of coverage; and insurers could have access to data allowing them to make more rational and timely coverage and reimbursement decisions.

Technology assessment is a classic illustration of free-market failure—i.e., market forces will not compel the generation of the optimal amount of performance data essential for the effective and efficient running of the system. Thus, the responsibility for assessing medical technologies cannot be left to the unaided, competitive forces of the marketplace. Because they would receive direct financial benefits, both Federal and private insurers would be expected to provide financial support for IHCE's work. However, because they have vested interests in certain outcomes (e.g., results justifying decreased utilization or advancing lower cost alternatives), these insurers should not have exclusive control of IHCE's operation. In order to maintain its credibility, the organization would have to be governed through a system of checks and balances involving constituent groups.

Goals and Objectives

IHCE's goal would be to generate cost-effectiveness data with a strong emphasis on the measurement of outcomes of therapeutic intervention. These data are needed by medical professionals as a basis for making decisions and informing patients about their choices in medical care; they are needed by health care consumers who are increasingly expected to assume responsibility for their own health and to participate in therapeutic decisions; and they are required by the insurance industry in order to design rational health insurance plans. Adequate technology assessment represents the core of the two major issues facing health care today: 1) how best to employ complex technologies, old as well as new, to meet the public's medical needs; and 2) how to limit the costs of medical care without jeopardizing its quality.

The proposed IHCE would have four major objectives:

- development of a uniform data base;
- systematic identification of agenda issues;
- generation of new data and analyses; and
- dissemination of information to carriers, professionals, and the public.

The achievement of the first objective, development of a uniform data base, is necessary to facilitate the collection of information from diverse sources. At present, a modest amount of relevant but often nonuniform data is generated by many health providers. For example, the Kaiser Foundation Health Plan, as part of patient registration, collects detailed information on service utilization at its various hospitals and clinics, then aggregates the information on a regional basis for planning of resource allocations. The proposed IHCE could develop guidelines for an instrument, e.g., a patient registration form, which could serve as both a receipt for billing and method for monitoring utilization and identifying new procedures. Decentralization of data storage at a regional, local, or even health plan level would help address the need for confidentiality of the sources of information. Each plan, locality, etc., might store its own standardized data and make these data available to IHCE-authorized researchers.

---

*Reference citations for this appendix on p. 190.

The second objective of the proposed IHCE, to serve as a communication clearinghouse that would systematically collect priority issues from its members, is an objective that has not been satisfactorily met in the current system, which relies primarily on signals received from claims data. IHCE would have the authority and capability for routinely surveying professionals to elicit their opinions about the future directions of innovations in their specialities. This model is currently being used by Kaiser-Oakland's Technology Assessment Division. Innovations and medical problem areas could also be identified in a number of other ways, including systematic literature reviews and monitoring of professional meetings. The patient registration forms could act as one "flagging" device. This task might be under the purview of existing and surviving Professional Standards Review Organizations (PSROs). Finally, health plans, private enterprise, and others might directly identify a new procedure and request evaluation by IHCE.

IHCE's third objective would be to participate in the generation of new data, specifically, through the support of clinical trials, retrospective studies, data banks, and possibly surveillance. Currently, a small number of clinical trials are funded primarily by NIH, but it can be anticipated that the NIH investment in this function will diminish in the near future (4), and currently proposed legislation specifically enjoins NCHSR from funding "clinical" studies. Clinical trials, although complex, time consuming, and resource intensive, remain the best method for determining the relative value of alternative medical technologies. Rather than being retreated from, clinical trials should be used more extensively within the limitations of their methodology and the clinical circumstances.

The fourth objective of IHCE could be a second function of its communication clearinghouse: dissemination of results of analysis back to the participants. Such dissemination could be achieved though the provision of access to computerized information, which could include annual reports, strategic objectives, findings of prior assessments, and listings of studies presently in progress. Participants would then be able to incorporate these results as they saw fit into their respective decisions. The proposed IHCE would *not* have responsibilities in policymaking. Its responsibilities would rest solely in the areas of data collection and analysis.

Funding

The perception of a need for an IHCE is based on the recognition that health care assessment is a public good. The marginal cost of assessment for any individual or group generally far exceeds the marginal benefit derived for any individual or group. The generation of information as a public good creates a "free-rider" problem. After an individual or group pays to determine that a particular procedure, protocol, device, etc., is more cost effective, the very act of capitalizing on that information makes it public knowledge; the information is freely available to other individuals and groups who did not share the cost of making the determination. The ethics of medical care encourage the early and broad dissemination of information.

The proposed IHCE would be a nonprofit organization funded by a per capita assessment or levy to be received from all qualified health plans. The funding for IHCE could be established on either a mandatory or voluntary basis.

With a mandatory system, health plans (for-profit and nonprofit) would be required to support IHCE as a condition of their receiving recognition as a "qualified" health plan—and therefore becoming eligible to receive tax credits, vouchers, or Medicare payments. To prevent the problem of free-riders (i.e., competing insurance programs which gain access to information without paying for the costs of its generation), a fee would be required to support IHCE. Although there are increasing numbers of industries that self-insure for their employees' medical care, most still carry administrative contracts with private insurers (or claims administrators). Under a mandatory structure, the fee for health care evaluation would be a required component of this administrative contract.

Other service organizations have set up similar models of cooperative research. One model is the Electric Power Research Institute, which assumes responsibility for part of the research agenda of the electric power providers. Participants in the Electric Power Research Institute contribute to the support of the institute without resorting to taxation such as that proposed in the mandatory version of IHCE suggested above. However, their situation differs from that of health care in two significant respects. First, the electric utilities do not compete with one another; they are a regulated monopoly. Second, these utilities all have a cost-based price regulation, allowing them to pass the cost of membership directly on to the consumers.

A second model for cooperative research is the Health Effects Institute, a nonprofit organization established in 1980 to study the health effects of automotive emissions. This institute is funded jointly by grants from governmental and charitable services and by additional funds contributed by participating automobile manufacturers according to a formula developed by industry. The Health Effects Institute is an independent organization that has no actual governance link to the Environmental Protection Agency, automotive industry, or public participants. Its health research committee establishes research priorities, develops research programs and protocols, obtains exhaust samples from manufacturers, and contracts with research centers to perform specified tasks.

Yet another example of voluntary cooperative research comes from the insurance industry itself. The Insurance Institute for Highway Safety, founded in 1959, is a nonprofit corporation established to study the contributory factors of drivers, vehicle design, and roadways to highway safety. The Insurance Institute's annual budget is based on contributions from three automotive insurance trade associations and one associated insurance group (whose members do not belong to any of the trade associations). These groups, in turn, raise their contributions from individual companies on the basis of their total premiums. The research protocols are developed by the Insurance Institute and contracted out to academic research centers. These voluntary models might not be feasible in an increasingly competitive health care environment, where some carriers could gain access to the information without paying for it and, thus, offer lower rates to subscribers than could carriers who were contributing members of IHCE.

Nevertheless, it might be possible to fund IHCE through voluntary contributions. Although this would create the free-rider problem described above, it might also alleviate some of the initial resistance to the establishment of such an institute. With voluntary funding, there would still be a membership fee required of all participants, which would cover IHCE's basic administrative costs. IHCE's governing body would develop an agenda of research topic alternatives on which the members of IHCE would vote. Alternatives would be given priorities by the membership, and members would subscribe in advance to cover the costs of conducting specific research studies.

What it would cost to develop the proposed IHCE is of obvious concern to those who would be expected to bear the burden of expenses. Relman has suggested that two-tenths of 1 percent (0.002) of expenditures for medical care might be an appropriate sum to allocate for this purpose (9). Since the current expenditures for medical care of private insurance and HCFA are approximately $160 billion and $85 billion, respectively, this would amount to nearly $500 million. Whether this is more or less than the task will require is by no means clear. If IHCE succeeds in its mission, this will be a small price to have paid. Indeed, as some suggest, the potential savings that could be expected to accrue as a result of better data on cost effectiveness would be many times greater than this amount (3,9).

IHCE's success, however, cannot be guaranteed. Therefore, rather than making an all-to-nothing commitment to a program of this magnitude, it would be prudent to proceed in modest steps. To test the proposed IHCE's potential, its board or council might begin by identifying those areas of medical care deemed to be in greatest need of evaluation, raising funds by assessment as needed for each subject of inquiry or analysis. Indeed, this would appear to be an appropriate method for the future funding of all projects: funds being generated only as the potential users of information judge necessary and appropriate.

Mechanism and Structure

IHCE could select topics for evaluation from those generated by its technology surveys. The topics would be given priority by the appropriate committee (board of directors or council). IHCE would let out contracts for clinical trials, retrospective studies, and technology assessments to carriers, as well as contracts for data analysis to professional organizations. In addition, IHCE would review and fund independently submitted proposals for clinical trials and for technology assessment. For example, investigators planning new procedures might apply directly to IHCE for funding, including the clinical costs of particular innovations. IHCE would coordinate its activities with those of other research organizations such as the disease-specific private foundations and Federal research agencies such as NCHSR, NCHCT (if funded), NIH, and the Food and Drug Administration.

It is anticipated that IHCE would be able to use some already established mechanisms for data collection, including claims data, health systems agencies (HSAs), and PSROs. HSAs and PSROs may succumb to current budget cuts, but the evaluation capabilities developed by the more successful PSROs could profitably be put to work on contract by IHCE (8).

Although the agenda would be set as a consensus at the national level, most of its implementation would take place under local control and responsibility. Individual institutions, through their institutional review boards, would determine whether proposed new or experimental procedures are used with appropriate standards for patient safety and whether standards for informed consent have been met.

The proposed IHCE would be governed by a board of directors or council composed of representatives from member groups: private insurance carriers, governmental insurers, HMOs, professional associations, and consumers. In addition to topical subgroups, there would be specific departments for legal affairs, communication, and publications.

Concerns With the Model

Legal Problems

Funding Sources.—If funding for IHCE were mandated by a tax, levy, or assessment to be paid by all insurers according to the number of individuals they cover, it would require new legislation. A Federal tax would presumably violate the current position of the insurance industry, which has been exempt from Federal legislation under the McCarran-Ferguson Act. Thus, taxation on private carriers would represent a substantial departure from this position. A tax on nonprofit organizations, including Blue Cross and Blue Shield and HMOs, would present similar problems. However, such a recommendation has a precedent in a recent proposal by David Stockman, Director of the Office of Management and Budget. This proposal, the Gephart-Stockman National Health Care Reform Act, would levy a tax on health insurers to provide funds for insuring subscribers against the financial failure of their selected health plans. The alternative of voluntary funding does not present these legal problems.

Research Authority.—Current insurance trends are generally to avoid involvement in research except under certain explicit circumstances. Therefore, it is uncertain on what legal grounds insurers would participate. Some legal advice indicates that insurers do have a research authority, but this concept is not explicit nor does it appear to have been tested in the courts.

Antitrust.—Various insurers have suggested that the operation of IHCE might be in violation of the Sherman Antitrust Act. However, it should be noted that many commercial carriers, through the Health Insurance Association of America, already jointly use the coverage recommendations of the Council of Medical Specialty Societies based on its program for clinical procedure review. The program for clinical procedure review identifies clinical procedures that are obsolete, duplicative, or not yet clinically proven. The Council of Medical Specialty Societies then recommends continuation or withdrawal of reimbursement. The proposed IHCE would differ from the Council of Medical Specialty Societies in that its function would be limited to the development, analysis, and distribution of data and would not include policy recommendations.

Selective Coverage.—The need for selective coverage has been widely recognized by third-party payers. To refuse payment to a hospital on the basis of inadequate experience, equipment, or trained staff, or failure to adhere to published standards, can be expected to lead to litigation—and has already done so. Part of the difficulty results from the wording of current insurance policies and contracts. Future policies should be drawn up explicitly indicating that coverage for specified procedures will be limited to certain providers. (Blue Shield of California is currently exploring the feasibility of contracting with better qualified hospitals and physicians for heart surgery at set prices substantially lower than standard or average fees; this is similar to the arrangement of preferred providers suggested for HMOs by Interstudy (7).) The possible need for legislation to protect third-party payers under such arrangements is deserving of exploration.

Selective coverage for new and experimental medical and surgical procedures might be provided by commercial carriers or by HCFA in conjunction with the proposed IHCE's program, with hospitals and physicians selected for participation in clinical trials on a case-by-case basis. Reimbursement under these conditions would represent what amounts to a research award and would presumably present less of a threat of litigation by nonparticipating and unapproved hospitals and physicians or their patients.

A serious problem that would remain, however, would be the opportunity for those physicians and/or hospitals not selected for reimbursement for a new (or old) procedure to do it anyway and to submit charges using billing codes for other, standard, procedures. This is one of the difficulties encountered in the current system of reimbursement.

Comprehensive restructuring of the method of reimbursement (e.g., capitation, prepayment, or payment by voucher) would partially resolve this problem, since a fixed amount of money would be available for all procedures. Additional funds for new procedures would have to be negotiated on a procedure-by-procedure basis. Short of such radical changes in method of reimbursement, natural forces within the present system may be expected to exert some, perhaps considerable, corrective influence. Malpractice suits against physicians who deviate from established standards are occurring with greater frequency. With increasing professional consensus that it is poor practice to perform specified complex procedures on an occasional basis or in facilities not equipped or staffed for such procedures, it can be anticipated that lawsuits will be brought against physicians and institutions that do this. Physicians will be forced to be more circumspect, and hospitals will look more and more to their institutional review boards for guidance in undertaking new and experimental procedures.

Ethical Issues

A fundamental principle of justice that has often been urged is that innovations of established efficacy be available to all of the population. This principle was most dramatically implemented by the Medicare Amendments of 1972, by which entitlement to medical care for end-stage renal disease (hemodialysis, kidney transplant) was conferred on all citizens of the United States. This principle has been repeatedly invoked in policy analyses of the artificial heart program (1), and it underlies current deliberations concerning reimbursement for heart transplants.

Heart transplantation in the United States has been concentrated primarily at Stanford University, where it is now considered an established clinical procedure with survival results comparable to those achieved with cadaver kidney transplants. Funded originally through an NIH research grant, the clinical costs of heart transplantation at Stanford were reimbursed by HCFA from 1979 until early 1981, and many private insurance companies now reimburse for heart transplantation. Relatively few heart transplants have been performed at other institutions, and their combined results have not matched Stanford's. Reimbursement by HCFA has not been made available to these other institutions because of their less satisfactory results. When challenged on this apparent inequity by another institution, HCFA responded by withdrawing reimbursement for heart transplantation at Stanford and announcing plans for a 2-year study of ethical, legal, and economic aspects of heart transplants. This is where the issue now stands.

When a medical or surgical procedure is clearly experimental, there can be no ethical obligation to make such a procedure available to all. An experimental procedure is, by definition, of unknown benefit. It may be better, or worse, or equal to previously available, established therapies or to no treatment at all. It is because of these procedures' unknown efficacy that mechanisms such as informed consent and institutional review committees have been established to advise prospective patients of the risks of such procedures and to reduce the possibilities of harm.

Approval for the performance of experimental procedures must rest on the qualifications of the investigators and on research resources and priorities, not solely on the medical needs of the patients or the availability of reimbursement from the insurers. Local institutional review committees are the appropriate agents to determine whether a proposed research procedure is scientifically and ethically justified and whether the interests of the patient, as experimental subject, are adequately protected. There are clear guidelines for these committees to follow in making these determinations.

The just distribution of efficacious medical care, in accordance with the foregoing general principles, requires better data than the data currently available. It is the exception, rather than the rule, that new therapies are introduced with well-controlled clinical trials leading to definitive evidence of therapeutic worth. In the absence of such data, effective treatments may be withheld or ineffective treatment may be given. Both errors seem likely to occur, in view of the many variations in procedure and hospitalization rates reported by Wennberg and others (12). Both errors, to the extent that they are avoidable, may be considered serious injustices; it is not clear that one is more serious than the other. Neither is it clear which error is the more common. However, there is strong presumptive evidence that when efficacy data are absent, physician-investigators tend to err in the direction of overestimating the potential benefits of therapy (5,6) and that many "unnecessary" procedures are carried out as a result of professional enthusiasm or optimism. The argument that there is widespread overprescribing of therapy has been developed in detail elsewhere (2). Overutilization of unproven medical interventions has immediate and urgent implications for distributive justice. Soon, society will no longer be able or willing to pay for all treatments that might be effective. Purchase of care that is ineffective or of undocumented efficacy for some patients will almost certainly result in the failure to provide effective care to other patients.

Quality of Information

A final and important concern relates to the quality of information to be collected by IHCE. Towery and Perry, at NCHCT, have proposed that "third-party payers, including Medicare . . . make reimbursement to providers contingent on their submitting certain minimal data under a previously agreed on protocol" (11). Sherman, Fineberg, and Frazier, at the Harvard School of Public Health, have made a similar proposal and have identified a number of contingencies for reimbursement (10). These proposals address the problem from the perspective of an agency whose principal responsibility is to provide reimbursement, and for whom the collection of data is by definition a secondary priority. The mandatory submission of data by medical care providers can also be assumed to be a secondary priority, and the quality of the resulting information may be poor.

The primary responsibility of the proposed IHCE, in contrast, would be to collect reliable information. Grants and contracts would be awarded on the basis of the anticipated quality of that information. Reimbursement for clinical services might be provided in conjunction with the grant or contract, but it would be a secondary consideration.

Discussion and Conclusions

The absence of a consistent and explicit policy of reimbursement for new technologies results in the escalation of charges and at the same time provides incentives to conceal innovations. In addition, there is no single organized and adequately funded program or agency charged with the responsibility for the generation of data with which to evaluate new (and old) technologies. As a result, not only are good outcome data lacking; often it is not even possible to identify when new procedures are performed. Indeed, the current system provides incentives which actively discourage the explicit identification of new and experimental technologies. An additional difficulty is the inability to provide reimbursement for these procedures on a selective basis. This inability makes it impossible to achieve the orderly development and evaluation of new technologies.

Potential remedies for the foregoing difficulties are readily at hand. At the regional and local levels, Medicare contractors, such as Blue Shield of California through its Medical Policy Committee, are already reviewing claims for new therapies. Unproven therapies are currently rejected for coverage, but could, having been identified, be selected for coverage contingent on collection of appropriate evaluation data and/or presentation of an appropriate experimental design.

To justify such a major shift in coverage policy would require major new funding for the evaluative process. There are, at present, instances where funding for clinical procedures and their evaluation is provided on an individual basis, at least in part, by the third-party payer. For example, Blue Shield of Massachusetts has paid the clinical costs of three diagnostic examinations for tumors of the adrenal, kidneys, and pancreas (ultrasound, computed tomography scan, and radionucleotides), the costs of data collection and analysis being funded by a foundation source. Blue Shield of California is funding an analysis of the cost effectiveness of ambulatory surgery. But even such modest efforts severely extend the fiscal capacity of individual insurers—and, at a time of intense competition for subscribers, tend to worsen rather than improve their competitive position. (The individual insurer must charge its subscribers the extra cost, but all insurers can use the information resulting from the investigation.) A resolution of this dilemma might be for the insurers to join in common purpose and to create a joint fund, from which amounts could be awarded by contract for proposals to evaluate specific procedures.

The basic provisions of this option can be summarized as follows: First, an IHCE would be created under the control of: 1) third-party payers, including Blue Shield, Blue Cross, and the commercial health insurance companies; 2) HMOs, represented by the Group Health Association of America and the American Association of Foundations of Medical Care; 3) the Government, represented by HCFA and (if funded) NCHCT; 4) the medical profession, represented by the Council of Medical Specialty Societies and/or the American Medical Association Council on Scientific Affairs; and 5) representatives of the public at large (consumers). Second, IHCE's goals would be: 1) the establishment of a uniform data base; 2) the systematic identification of agenda issues; 3) the generation of new data and analysis; and 4) the dissemination of information to carriers, professionals, and consumers. Third, IHCE would be funded through fees or contributions from public and private insurers (including self-insurers) and/or from HMOs on either a mandatory or voluntary basis. Finally, new and experimental medical and surgical procedures would be selectively covered on the basis of locally approved research protocols and the availability of data for independent analysis.

Appendix F References

1. Artificial Heart Assessment Panel of the National Heart and Lung Institute, *The Totally Implantable Artificial Heart: Legal, Social, Ethical, Medical, Economic, Psychological Implications,* DHEW publication No. (NIH) 74-191, June 1973.
2. Bunker, J. P., "Health Care Costs," in *Encyclopedia Britanica, Medical and Health Annual,* in press, 1982.
3. Bunker, J. P., "Hard Times for the National Centers," *N. Eng. J. Med.* 303:580, 1980.
4. Frederickson, D. S., "Biomedical Research in the 1980's," *N. Eng. J. Med.* 304:509, 1981.
5. Gilbert, J. P., et al., "Statistics and Ethics in Surgery and Anesthesia," *Science* 198:684, 1977.

6. Grace, N. D., et al., "The Present Status of Shunts for Portal Hypertension in Cirrhosis," *Gastroenterology* 50:684, 1966.
7. Interstudy, Minneapolis, Minn., "Preferred Provider Arrangements-California," internal memorandum to P. Ellwood from K. Lewis, Nov. 18, 1980.
8. Lohr, K. N., et al., *Professional Standards Review Organizations and Technology Assessment in Medicine*, prepared by the Rand Corp. for the Office of Technology Assessment, U.S. Congress, Washington, D.C., March 1981.
9. Relman, A. S., "Assessment of Medical Practices," *N. Eng. J. Med.* 303:153, 1980.
10. Sherman, H., et al., "Issues Raised by Third Party Reimbursement for Investigational Procedures: Options and Criteria," report to the National Center for Health Care Technology, Feb. 2, 1980.
11. Towery, O. B., and Perry, S., "The Scientific Basis for Coverage Decisions by Third-Pary Payers," *J.A.M.A.* 245:59, 1981.
12. Wennberg, J. E., and Gittelsohn, A., "Small Area Variations in Health Care Delivery," *Science* 182:1102, 1973.

Appendix G.—Method of the Study and Description of Other Volumes

Method of the Study

The study *Strategies for Medical Technology Assessment* began on July 1, 1980. Immediately thereafter, a planning period was begun, and an advisory panel was selected.

Most of the studies undertaken at OTA rely on the advice and assistance of an advisory panel of experts. The advisory panel for a particular assessment suggests source materials, subject areas, case studies, and perspectives to consider; assists in interpreting information and points of view that are assembled by OTA staff; and suggests possible findings and conclusions based on the accumulation of information produced by the study. The panel members review staff and contract materials for accuracy and validity, discuss policy options of the study, and present arguments for and against the options and conclusions. However, they do not determine the report's final form and are not responsible for its content, direction, or conclusions.

The advisory panel for this assessment consisted of 19 men and women with backgrounds in medicine, public health, sociology, information and library science, economics. law, psychiatry, consumer advocacy, technology assessment, industry, health policy, ethics, and health insurance. The panel was chaired by Lester Breslow of the University of California at Los Angeles. One member of the OTA Health Program Advisory Committee, Kerr White, also served on the panel.

The first panel meeting was held on September 12, 1980, in Washington, D.C. (the site of all three panel meetings). Prior to the meeting, panel members were sent a detailed study plan, including a suggested outline, and several pertinent articles as background for discussion. During the meeting, panel members discussed the overall study plan for the assessment and helped OTA staff refine the goals for the project. The panel examined the project boundaries and definitional issues and was key in sharpening the study's focus. The panel was also helpful in reviewing the primary issue areas to be covered and in providing suggestions of individuals and organizations to contact for information and assistance. The panel was particularly helpful in suggesting modifications in several of the contractors' reports (which were just beginning). Several contractors were present and participated in the meeting.

By the fall of 1980, all of the major contracts for the main report were let. Each contract effort is described below:

- John Williamson (Johns Hopkins University) developed a helpful and imaginative theoretical framework for a strategy for medical technology assessment.
- Kathleen Lohr and Robert Brook (Rand Corp.) were asked to explore the potential role of the Professional Standards Review Organizations (PSROs) in a medical technology assessment system. With the assistance of John Winkler (Rand), they examined how physicians received new medical knowledge, the potential for PSROs to transfer that knowledge, and the ability of PSROs to test technologies for safety and effectiveness. Their work is available from Rand as a published document.
- Paul Wortman (University of Michigan) and Leonard Saxe (Boston University) analyzed the methods of testing medical technologies, synthesizing information and soliciting group opinions. Their paper, reproduced as appendix C, formed the basis for much of chapters 3 and 5 of this report.
- John Reiss (Baker & Hostetler) was asked to write a paper on the role of reimbursement policy in an assessment strategy. His ideas are included in many sections of this report.
- John Bunker collaborated with Jinnet Fowles (both of Stanford) and a number associates to write a paper on the effects of reimbursement on the innovation process for medical and surgical procedures. They developed the concept for the "Institute for Health Care Evaluation," reproduced as appendix F, and submitted several case studies, two of which are included in appendix E. Their main work is published as a two-part series by the *New England Journal of Medicine* (Mar. 4 & 11, 1982).
- John Wennberg (Dartmouth) submitted an interesting paper on the use of health insurance claims and patient outcome data to evaluate health care technologies after they are generally available.
- Patricia Woolf (Princeton) prepared a paper describing private sector activities in both bibliographic data base production and vending. Appendix B contains a listing of those bibliographic-related resources.

Contractors' reports were reviewed both by OTA staff and by a large number of outside experts. Reviewer's comments were forwarded to the authors, who incorporated them in revising drafts.

In January 1981, a workshop was held in Boston to review the first draft of the Wortman/Saxe paper on the role of various methods for testing medical technologies for safety, efficacy, and effectiveness. Participants included the authors and other scientists from several research disciplines, an advisory panel member, and OTA staff. Wortman and Saxe prepared a second draft of their paper on methods based largely on that workshop.

The second advisory panel meeting was held on January 28, 1981. The discussion of that meeting centered around three main topics: 1) the selection of methods for testing medical technologies, 2) the use of the PSROs in technology assessment, and 3) the Medical Literature Analysis and Retrieval System (MEDLARS) and the National Library of Medicine (NLM). Panel comments were helpful in all three areas of study and were instrumental in determining the role of all three areas in an assessment strategy.

OTA staff produced two staff papers for congressional hearings held by the House Energy and Commerce Committee and the Senate Labor and Human Resources Committee in March 1981. One paper dealt with NLM, the other with the National Center for Health Care Technology (NCHCT). After the hearings, both papers were reviewed widely and were subsequently incorporated in this report and related technical memorandum.

A workshop on reimbursement policies and innovation met in Washington on May 13, 1981. The purpose of this workshop was to review the papers by Reiss and by Bunker and generally to discuss the effects that medical technology assessment and reimbursement policy have on innovation. In addition to the authors and OTA staff, workshop participants included several members from the study advisory panel, an inventor, representatives from industry and researchers and academicians interested in medical technology innovation. The workshop was helpful to Reiss and Bunker in revising their papers and was particularly helpful to OTA in understanding both the innovation process and the effects that an assessment policy may have on that process.

While the main project was proceeding, several subcomponents began to take on added significance as separate projects in their own right. Following the March Senate Labor and Human Resources Committee hearing, OTA was asked by Chairman Hatch to prepare a separate technical memorandum by expanding the NLM component of the study to examine the role of both the private and public sectors in producing and vending bibliographic data bases. In addition, a separate volume concerning the postmarketing surveillance of drugs was prepared in draft. Subsequently, OTA was asked by the House Committee on Energy and Commerce to elevate the postmarketing surveillance effort to full report status (i.e., complete with options). It is being published as volume II of this assessment. And finally, the House Committee on Energy and Commerce also asked OTA to write a separate report on the effects which competitive health care system proposals may have on medical technologies. The Senate Committee on Labor and Human Resources endorsed that request.

Prior to the third panel meeting, held on September 22, 1981, an initial draft of the final report was prepared and sent to panel members. The entire meeting was spent reviewing that draft and focused primarily on the policy analysis and options for Congress.

The draft was then revised by OTA staff on the basis of the suggestions and comments of the advisory panel. The revised draft was then sent for a further round of review by a much broader range of experts in a diversity of settings: Federal agencies, private and nonprofit organizations, academic institutions, practicing health professionals, consumer groups, and other selected individuals. Altogether, more than 150 individuals or organizations were asked to comment on drafts of the main volume or other components of this assessment. The main volume, containing policy options, was reviewed by approximately 100. Following revisions by OTA staff, the report was submitted to the Technology Assessment Board.

Description of Other Volumes

This assessment has resulted in seven documents:

1. the main report, of which this index is a part;
2. a brochure that summarizes the main report;
3. a staff paper on NCHCT, issued to Congress in March 1981;
4. a staff paper on NLM, also issued to Congress in March 1981;
5. a monograph published by Rand Corp. entitled *Peer Review and Technology Assessment in Medicine;*
6. a full report, which is volume II of this assessment, entitled *Postmarketing Surveillance of Prescription Drugs;* and
7. a technical memorandum on *MEDLARS and Health Information Policy.*

Brief descriptions of the last two volumes are provided below. Also described below is a volume entitled *Medical Technology Under Proposals To Increase Competition in Health Care.* This report grew out of the *Strategies* assessment but is now being published separately.

Postmarketing Surveillance of Prescription Drugs (Vol. II of *Strategies for Medical Technology Assessment*)

To market a drug, manufacturers must provide evidence of its efficacy and safety to the U.S. Food and Drug Administration (FDA). Once these premarketing requirements are met and a drug is released, FDA can suggest but cannot impose restrictions on the drug's use. However, it can remove the drug from the market for reasons such as new evidence on safety or efficacy, any untrue statement of a material fact, or failure to meet manufacturing standards.

In the premarketing clinical tests, controlled clinical trials with limited numbers of test subjects are used. Observation of limited numbers of patients for a short period of time uncovers minimal information about a drug's potential uses and dangers, and postmarketing activities of various types have been proposed over the past decade. However, postmarketing surveillance has been linked to other policy objectives, such as speeding up the premarketing approval process and using postmarketing information to improve physician prescribing practices.

In OTA's report, the analysis of these issues is framed around the following questions: 1) what aspects of the premarketing requirements might be curtailed to shorten the drug approval process? 2) what additional powers would help strengthen FDA's activities in the postmarketing period? and 3) what possible tradeoffs might there be between curtailing some premarketing requirements and strengthening FDA's role in monitoring drugs once they are released into the marketplace?

MEDLARS and Health Information Policy

This technical memorandum examines the role of NLM's computerized bibliographic retrieval and technical processing system, MEDLARS, with respect to the role of private sector information systems in the creation and distribution of computerized bibliographic health-related information.

The study examines two specific sets of issues: the range of NLM's computerized products and services, and NLM's pricing structure for leasing data base tapes and for online access to the data bases. It focuses on the domestic and international implications of these issues and stresses the importance of new and emerging computer and communications technologies on biomedical information policy.

The issues are considered within a general framework of the Government's role in the allocation of resources to information development and distribution, the effect of the Government's involvement in allocative activities on certain segments of the private information sector and the health community, and the historic role of the Government in health information activities.

Medical Technology Under Proposals To Increase Competition in Health Care

Proposals to stimulate competition in medical care tend to fall into three categories: 1) increased cost sharing by patients for services, 2) increased competition and hence greater pressure for efficiency among health plans or providers, and 3) increased antitrust activities. OTA's study considers only the first two.

The study focuses on the effects and policy implications of the different proposals in three major areas of medical care: 1) consumer information, 2) quality of care, and 3) technology innovation and use. In each of these areas, the study examines the situation that would pertain under each type of proposal and any differences from the present situation, effects that could be expected, any problems that would arise, and methods of addressing these problems.

Appendix H.—Acknowledgments

OTA Contractors

The following individuals participated in preparing background papers or other material for this study. Many of them also provided valuable advice and comments during the course of the assessment.

Herbert Abrams
Stanford University

Robert H. Brook
Rand Corp.

Alain C. Enthoven
Stanford University

Renée Ford
Tifford Producers, Ltd.

Jinnet Fowles
Stanford University

Peter G. Goldschmidt
Veterans Administration

H. Diane Goroum
Stanford University

Katherine Jones
Stanford University

Aida A. LeRoy
Health Information Designs

Kathleen N. Lohr
Rand Corp.

Barbara S. Lynch
Rockville, Md.

Margaret Marnell
Stanford University

M. Lee Morse
Health Information Designs

David Sawi
Stanford University

Ralph Schaffanzick
Blue Shield, North Carolina

Anne A. Scitovsky
Palo Alto Research Foundation

John E. Wennberg
Dartmouth College

John D. Winkler
Rand Corp.

Patricia Woolf
Princeton University

Study Workshops

Two helpful workshops were held during the course of this assessment, one on methodological issues, the other on reimbursement and innovation issues. OTA appreciates the participation of the following people, many of whom also provided thoughtful comments on drafts of papers and of this report.

Robert Austrian
University of Pennsylvania

David Blumenthal
Harvard University

Lester Breslow
University of California, Los Angeles

John P. Bunker
Stanford University

Ralph D'Agostino
Boston University

D. V. d'Arbeloff
Millipore Corp.

Harvey Fineberg
Harvard University

Stanley B. Jones
Blue Cross/Blue Shield

Alan Kahn
Madison, Wis.

John Kimberly
Yale University

Aida A. LeRoy
Health Information Designs

M. Lee Morse
Health Information Designs

Vernon L. Nickel
Sharp Rehabilitation Center

John B. Reiss
Baker & Hostetler

Richard A. Rettig
Rand Corp.

Louise B. Russell
The Brookings Institution

Leonard Saxe
Boston University

Alexander M. Walker
Harvard University

Richard N. Watkins
Group Health Cooperative

John E. Wennberg
Dartmouth Medical School

John W. Williamson
The Johns Hopkins University

Paul M. Wortman
University of Michigan

Reviewers

A large number of individuals provided valuable advice and comments to OTA during this assessment. OTA is grateful for their assistance. In particular, we would like to thank the following people:

John C. Bailor, III
Environmental Protection Agency

Constance Barton
National Center for Health Care Technology

Dwight Blankenbaker
National Center for Health Care Technology

Queta Bond
Institute of Medicine

John T. Bruer
Rockefeller Foundation

Arthur Caplan
Hastings Center

Gerald Cohen
National Center for Health Services Research

Thi D. Dao
Pharmaceutical Manufacturers Association

Carolyne K. Davis
Health Care Financing Administration

W. Palmer Dearing
Blue Cross/Blue Shield

Robert A. Derzon
Lewin & Associates

Manning Feinleib
National Institutes of Health

Deborah A. Freund
University of North Carolina at Chapel Hill

Sydney P. Galloway
Health Care Financing Administration

Stephen F. Gibbons
Department of Health and Human Services

Donald Goldstone
National Center for Health Services Research

Roger Graham
Blue Cross/Blue Shield

Nancy Greenspan
Health Care Financing Administration

Ruth S. Hanft
Department of Health and Human Services

Dorcas R. Hardy
Department of Health and Human Services

Jeffrey E. Harris
Massachusetts Institute of Technology

Itzhak Jacoby
National Institutes of Health

Judith K. Jones
Food and Drug Administration

John T. Kalberer, Jr.
National Institutes of Health

James M. Kaple
Health Care Financing Administration

Joel Kavet
Department of Health and Human Services

James R. Kimmey
Institute for Health Planning

Daniel Koretz
Congressional Budget Office

Sam Korper
Department of Health and Human Services

Peter Kramer
Alcohol, Drug Abuse, and Mental Health Administration

Marvin Kristein
National Institutes of Health

Charles U. Lowe
National Institutes of Health

David L. Martin
Health and Welfare, Canada

Alan E. Mayers
National Center for Health Services Research

David McCallum
National Institues of Health

Mary Louise McIntyre
Health Care Financing Administration

H. Calvin Meadows
Health Resources Administration

Peggy A. Miller
Kaye, Scholer, Fierman, Hays & Handler

Anthony T. Mott
Finger Lakes Health Systems Agency

Donald Murphy
National Institutes of Health

Donald Muse
Health Care Financing Administration

Beverlee A. Myers
California Department of Health Services

Seymour Perry
National Center for Health Care Technology

Harold A. Pincus
National Institute of Mental Health

Pierre Renault
National Center for Health Care Technology

John Rodak
National Center for Health Care Technology

Wayne I. Roe
Health Industry Manufacturers Association

Lawrence Rose
National Center for Health Services Research

Gerald Rosenthal
Public Health Service

Theodore Ryerson
National Technical Information Service

Linda Schaeffer
National Center for Health Care Technology

Henry Schoolman
National Library of Medicine

Stuart O. Schweitzer
University of California, Los Angeles

Ira D. Singer
Project Hope

Kent A. Smith
National Library of Medicine

Helen Smits
Yale University

Edward J. Sondik
National Institutes of Health

Barbara Starfield
The Johns Hopkins University

Michael A. Stoto
Harvard University

Stephen M. Stowe
University of Southern California

Paul R. Torrens
University of California, Los Angeles

James R. Ullom
National Center for Health Services Research

Donna Valtri
Office of Technology Assessment

Sally W. Vernon
University of Texas, Houston

Judith L. Wagner
Urban Institute

Monica Walters
National Institutes of Health

John M. Weinger
University of Southern California

Victoria Weisfeld
Institute of Medicine

Norman Weissman
National Center for Health Care Technology

Elizabeth Wenchel
Arthur Young & Co.

Daniel Wikler
University of Wisconsin, Madison

Irwin Wolkstein
Health Policy Alternatives, Inc.

Daniel I. Zwick
Bethesda, Md.

Appendix I.—Glossary of Acronyms and Terms

Glossary of Acronyms

ADAMHA	Alcohol, Drug Abuse, and Mental Health Administration (PHS)
AHA	American Hospital Association
AKCUP	Artificial Kidney-Chronic Uremia program (NIH)
AMA	American Medical Association
ANDA	abbreviated new drug application
BC/BS	Blue Cross/Blue Shield
BHP	Bureau of Health Planning (HRA)
BLS	Bureau of Labor Statistics (DOL)
BRS	Bibliographic Retrieval Service
CABG	coronary artery bypass graft
CAPD	continuous ambulatory peritoneal dialysis
CAS	Chemical Abstracts Service
CAT	computerized axial tomography (scanner)
CBA	cost-benefit analysis
CBO	Congressional Budget Office (U.S. Congress)
CCU	coronary care unit
CDC	Centers for Disease Control (PHS)
CEA	cost-effectiveness analysis
CEA/CBA	cost-effectiveness analysis/cost-benefit analysis (when referred to as a class of analytical techniques)
CEAP	Clinical Efficacy Assessment Project
CHAMPUS	Civilian Health and Medical Program of the Uniformed Services
CHSS	Cooperative Health Statistics System
CON	certificate of need
CPHA	Commission on Professional and Hospital Activities
CT	computed tomography (scanner)
DHEW	Department of Health, Education, and Welfare (now DHHS)
DHHS	Department of Health and Human Services (formerly DHEW)
DIALOG	DIALOG Information Services, Inc.
DOD	Department of Defense
DOE	Department of Energy
DOL	Department of Labor
ECRI	Emergency Care Research Institute
EFM	electronic fetal monitoring
ESRD	end-stage renal disease
FDA	Food and Drug Administration (PHS)
GAO	General Accounting Office (U.S. Congress)
HCFA	Health Care Financing Administration (DHHS)
HEW	Department of Health, Education, and Welfare (now DHHS)
HHS	Department of Health and Human Services (formerly DHEW)
HMO	health maintenance organization
HRA	Health Resources Administration (PHS)
HSA	health systems agency
HSQB	Health Standards and Quality Bureau (HCFA)
HUP	Hospital Utilization Project
IDE	investigational device exemption
IHCE	Institute for Health Care Evaluation (proposed)
IND	investigational exemption for a new drug
IOM	Institute of Medicine (National Academy of Sciences)
IPD	intermittent peritoneal dialysis
KDCP	Kidney Disease Control program (PHS)
MEDLARS	Medical Literature Analysis and Retrieval System (NLM)
MEDLINE	MEDLARS On-line (NLM)
MSAFP	maternal serum alpha-fetoprotein
NAS	National Academy of Sciences
NASA	National Aeronautics and Space Administration
NCHCT	National Center for Health Care Technology (OASH)
NCHS	National Center for Health Statistics (OASH)
NCHSR	National Center for Health Services Research (OASH)
NCI	National Cancer Institute (NIH)
NDA	new drug application
NDI	national death index
NECGD	nonequivalent control group design
NEI	National Eye Institute (NIH)
NGT	nominal group technique
NHLBI	National Heart, Lung, and Blood Institute (NIH)
NIA	National Institute on Aging (NIH)
NIAAA	National Institute on Alcohol Abuse and Alcoholism (ADAMHA)
NIADDK	National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases (NIH)
NIAID	National Institute of Allergy and Infectious Diseases (NIH)
NIAMDD	National Institute of Arthritis, Metabolism, and Digestive Diseases (now NIADDK)

NICHD	National Institute of Child Health and Human Development (NIH)
NIDA	National Institute on Drug Abuse (ADAMHA)
NIDR	National Institute of Dental Research (NIH)
NIEHS	National Institute of Environmental Health Sciences (NIH)
NIGMS	National Institute of General Medical Sciences (NIH)
NIH	National Institutes of Health (PHS)
NIMH	National Institute of Mental Health (ADAMHA)
NINCDS	National Institute of Neurological and Communicative Disorders and Stroke (NIH)
NLM	National Library of Medicine (NIH)
NSF	National Science Foundation
NTIS	National Technical Information Service (Department of Commerce)
OASH	Office of the Assistant Secretary for Health (DHHS)
OASPE	Office of the Assistant Secretary for Planning and Education (DHHS)
OHRST	Office of Health Research, Statistics, and Technology (OASH)
OMAR	Office for Medical Applications of Research (NIH)
OPE	Office of Planning and Evaluation (OASH)
OPEL	Office of Planning, Evaluation, and Legislation
OPPR	Office of Policy Planning and Research (HCFA)
ORD	Office of Research and Demonstrations (HCFA)
OTA	Office of Technology Assessment (U.S. Congress)
OTC	over-the-counter (Drug Review program) (FDA)
PAS	Professional Activity Study
PHS	Public Health Service (DHHS)
PSRO	Professional Standards Review Organization, Office of PSRO (HCFA)
PTA	percutaneous transluminal angioplasty
PTCA	percutaneous transluminal coronary angioplasty
QALY	quality-adjusted life year
R&D	research and development
RCT	randomized clinical trial
SDC	System Development Corp.
SEER	Surveillance, Epidemiology, and End Results program (NCI)
SSA	Social Security Administration (DHHS)
TA	technology assessment
TCC	Technology Coordinating Committee (DHHS)
VA	Veterans Administration

Glossary of Terms

Biomedical and behavioral research: A combination of biological, medical, psychological, social, and physical scientific investigations focused on eradicating disease and generating new scientific knowledge.

Capitation financing method: The method of paying for medical care on a fixed, periodic prepayment basis per individual enrolled in a health plan. Payment by "capitation" implies that the amount paid by the individual is independent of the number of services that individual has received.

Case-control study: An observational study design, referred to by some authors as "retrospective," in which individuals with a condition of interest (e.g., a suspected adverse effect of a medical treatment), i.e., cases, are compared to individuals without the condition, i.e., controls, with respect to factors (e.g., previous exposure to the treatment) which are judged relevant.

Certificate of need (CON): A regulatory planning mechanism required by the National Health Planning Resources Development Act of 1974 to control large health care capital expenditures. Each State is required to enact a CON law. CON applications by institutions are reviewed by local health systems agencies, who recommend approval or disapproval; they are denied or approved by State health planning and development agencies.

Cohort study: An observational study design, referred to by some authors as "prospective," in which two (or more) groups who vary with respect to their exposure to a factor of interest (e.g., a treatment method) are observed over a period of time. The status of individuals in all groups is assessed after an appropriate interval, and the outcomes compared to determine the effect of the factor of interest.

Consensus development conference: A process in which biomedical researchers, practicing health professionals, and others, as appropriate, are brought together by the National Institutes of Health to explore publicly the scientific background, state of knowledge, proper use(s), and any other issues pertinent to the technology under consideration.

Cost-benefit analysis (CBA): An analytical technique that compares the costs of a project or technological application to the resultant benefits, with both costs and benefits expressed by the same measure. This measure is nearly always monetary.

Cost-effectiveness analysis (CEA): An analytical technique that compares the costs of a project or of alternative projects to the resultant benefits, with costs and benefits/effectiveness expressed by different measures. Costs are usually expressed in dollars, but benefits/effectiveness are ordinarily expressed in terms such as "lives saved," "disability avoided," "quality-adjusted life years saved," or any other relevant objectives. Also, when benefits/effectiveness are difficult to express in a common metric, the may be presented as an "array".

CEA/CBA: A composite term referring to a family of analytical techniques that are employed to compare costs and benefits of programs or technologies. Literally, the term as used in this assessment means "cost-effectiveness analysis/cost-benefit analysis."

Data base: An organized collection of information in machine-readable form and accessible by computer.

Device (medical): Any physical item, excluding drugs, used in medical care (including instruments, apparatus, machines, implants, and reagents).

Discounting: A procedure used in economic analysis to reduce to present value those costs and effects that occur in future years. Discounting is based on two premises: 1) individuals prefer to receive benefits today rather than in the future; and 2) resources invested today in alternative programs could earn a return over time.

Distributive justice: A philosophical concept whose objective is to ensure that benefits in society are allocated in proper proportion to each individual's legitimate claim to them.

Drug: Any chemical or biological substance that may be applied to, ingested by, or injected into humans in order to prevent, treat, or diagnose disease or other medical conditions.

Effectiveness: Same as efficacy (see below) except that it refers to " . . . average or actual conditions of use."

Efficacy: The probability of benefit to individuals in a defined population from a medical technology applied for a given medical problem under ideal conditions of use.

Epidemiology: The study of the distribution, determinants, and control of diseases in human populations.

Experimental method: Any method of hypothesis-testing in which the investigator controls the application or withholding of the factor under study to individuals (or animals). Clinical trials (with control groups) of all types fall into this category.

Fee-for-service: A method of paying for medical care on a retrospective basis by which each service actually received by an individual bears a related charge.

Health maintenance organization (HMO): A health care organization that acts as both insurer and provider of comprehensive but specified medical services by a defined set of physicians to a voluntarily enrolled population paying a prospective per capita fee (i.e., paying by "capitation").

Health services research: A field of inquiry that focuses on the structure, production, distribution, and effects of delivering personal health services.

Health systems agency (HSA): One of the approximately 200 local health planning agencies designated under the National Health Planning and Resources Development Act of 1974 to develop local health planning goals and implement plans in consonance with State and national health care goals. HSAs are federally funded and are governed by a body which is broadly representative of both provider and consumer interests, the latter being in the majority.

Incidence: In epidemiology, the number of cases of disease, infection, or some other event having their onset during a prescribed period of time in relation to the unit of population in which they occur. It measures morbidity or other events as they happen over a period of time.

Marginal benefit: An economic concept referring to the additional benefit achieved by incurring an additional unit of cost.

Marginal cost: An economic concept referring to the additional cost of achieving one more unit of benefit.

Medicaid: A Federal program that is administered and operated individually by each participating State government that provides medical benefits to certain low-income persons in need of health and medical care.

Medical technology: The drugs, devices, and medical and surgical procedures used in medical care, and the organizational and supportive systems within which such care is provided.

Medicare: A nationwide, federally administered health insurance program authorized in 1965 to cover the cost of hospitalization, medical care, and some related services for eligible persons over age 65, persons receiving Social Security Disability Insurance payments for 2 years, and persons with end-stage renal disease. Medicare consists of two separate but coordinated programs—hospital insurance (part A) and supplementary medical insurance (part B). Health insurance protection is available to insured persons without regard to income.

MEDLARS: The computerized Medical Literature Analysis and Retrieval System of the National Library of Medicine. Available through a network of centers at more than 1,000 universities, medical schools, hospitals, Government agencies, and commercial organizations, MEDLARS contains some 4,500,000 references to journal articles and books in the health sciences published after 1965.

MEDLINE (MEDLARS On-line): The National Library of Medicine's online data base containing approximately 600,000 references to biomedical journal articles published in the U.S. and 70 foreign countries in the current and preceding 2 years.

Morbidity: A measure of illness, injury, or disability in a defined population. It is usually expressed in general or specific rates of incidence or prevalence. Sometimes used to refer to any episode of disease. See also "mortality (death)."

Mortality (death): A measure of deaths, used to describe the relation of deaths to the population in which they occur. The mortality rate (death rate) expresses the number of deaths in a unit of population within a prescribed time.

Observational method: Any method of hypothesis-testing in which the investigator does not control the application or withholding of the factor under study to individuals (or animals).

On-line: A term applied to a computerized "interactive" information retrieval system that allows an information specialist (or other user) sitting at a remote processing facility (i.e., typewriter or video terminal) to engage in a direct dialog with a central computer on which information (e.g., data bases, indexes) is stored, and thus to have immediate access to that information. The central computer and the information stored on the computer are said to be on-line to the remote processing facility(ies).

Prevalence: In epidemiology, the number of cases or disease, infected persons, or persons with disabilities or some other condition, present at a particular time and in relation to the size of the population. It is a measure of morbidity at a point in time.

Procedure (medical or surgical): A medical technology involving any combination of drugs, devices, and provider skills and abilities. Appendectomy, for example, may involve at least drugs (for anesthesia), monitoring devices, surgical devices, and the skilled actions of physicians, nurses, and support staffs.

Professional Standards Review Organizations (PSROs): Community-based, physician-directed, nonprofit agencies established under the Social Security Amendments of 1972 to monitor the quality and appropriateness of institutional health care provided to Medicare and Medicaid beneficiaries.

Randomization: The assignment by an investigator of individuals to treatment or control groups based on chance alone.

Randomized clinical trial (RCT): An experimental design by which human or animal subjects are randomly assigned, either to an experimental group (in which subjects receive the treatment being studied) or to a control group (in which subjects do not receive the treatment being studied). Also referred to as "randomized controlled clinical trial" or "controlled clinical trial."

Reliability: A measure of the consistency of a method in producing results. A reliable test gives the same results when applied more than once under the same conditions. Also called "precision."

Risk: A measure of the probability of an adverse or untoward outcome and the severity of the resultant harm to health of individuals in a defined population associated with use of a medical technology applied for a given medical problem under specified conditions of use.

Risk-benefit analysis: The formal comparison of the probability and level of adverse or untoward outcomes v. positive outcomes for any given action. The comparison of outcomes does not take into consideration the resource costs involved in the intended action.

Safety: A judgment of the acceptability of risk (see above) in a specified situation.

Technology: The application of organized knowledge to practical ends.

Technology assessment: A comprehensive form of policy research that examines the technical, economic, and social consequences of technological applications. It is especially concerned with unintended, indirect, or delayed social impacts. In health policy, the term has also come to mean any form of policy analysis concerned with medical technology, especially the evaluation of efficacy and safety. The comprehensive form of technology assessment is then termed "comprehensive technology assessment."

Validity: A measure of the extent to which an observed situation reflects the "true" situation. *Internal validity* is a measure of the extent to which study results reflect the true relationship of a "risk factor" (e.g., treatment or technology) to the outcome of interest in study subjects. *External validity* is a measure of the extent to which study results can be generalized to the population which is represented by individuals in the study which assumes that the characteristics of that population are accurately specified.

References

References

(References for authored papers in apps. E and F are cited at the end of each paper.)

1. "Adjuvant Chemotherapy of Breast Cancer," National Institutes of Health Consensus-Development Conference Statement, *N. Eng. J. Med.* 303(14):831, 1980.
2. Almy, T., Sounding Board, "The Role of the Primary Physician in the Health-Care Industry," *N. Eng. J. Med.* 304(4):225, 1981.
3. Altman, S. H., and Blendon, R. (eds.), *Medical Technology: The Culprit Behind Health Care Costs,* proceedings of the 1977 Sun Valley Forum on National Health, publication No. (PHS) 79-3216 (Washington, D.C.: U.S. Government Printing Office, 1979).
4. American College of Physicians, Philadephia, Pa., "ACP News . . . ," Nov. 24, 1980.
5. American College of Physicians, Philadelphia, Pa., "Clinical Efficacy Assessment Project: A Proposal for Expansion," mimeo, n.d.
6. Applied Management Sciences, Inc., and Cuadra Associates, Inc., *Catalog of Human Services Information Resources Organizations,* prepared for the Office of the Assistant Secretary for Planning and Evaluation, Department of Health and Human Services, contract No. HEW 100-79-0183, June 1980; *Project Share Human Services Monograph No. 15,* October 1980.
7. Arnstein, S. R., "Technology Assessment: Opportunities and Obstacles," *IEEE Trans. Syst. Man & Cybern.* SMC-7:571, 1977.
8. Arthur Young & Co., and Emergency Care Research Institute, *A Profile of the Medical Technology Industry and Government Policies,* draft final report to the National Center for Health Services Research, Hyattsville, Md., Mar. 31, 1981.
9. Astvad, K., et al., "Mortality From Acute Myocardial Infarction Before and After Establishment of a Coronary Care Unit," *Br. Med. J.* 1:567, 1974.
10. Bader, M., letter to the editor, *N. Eng. J. Med.* 96(1):46, 1977.
11. Banta, H. D., "Biomedical Research and Technology Assessment in the Major Medical Clinics," presented at the Project HOPE Conference, "The Role of Major Medical Clinics in Meeting Future National Health Needs," Millwood, Va., Mar. 11-13, 1981.
12. Banta H. D., "The Diffusion of the Computed Tomography (CT) Scanner in the United States," *Int. J. Health Serv.* 10(2):251, 1981.
13. Banta, H. D., and Behney, C. J., "Medical Technology: Policy and Problems," *Health Care Mgt. Rev.* 5:45, 1980.
14. Banta, H. D., and Behney, C. J., "Policy Formulation and Technology Assessment," prepared for the Experiment and Social Policy Issue of Health and Society, *Milbank Mem. Fund Q.* 59(3):445, 1981.
15. Banta, H. D., and McNeil, B. J., "Evaluation of the CAT Scanner and Other Diagnostic Technologies," *Health Care Mgt. Rev.* 3:7, 1978.
16. Banta, H. D., and Sanes, J. R., "Assessing the Social Impacts of Medical Technologies," *J. Comm. Health* 3(3):245, 1978.
17. Banta, H. D., and Thacker, S. B., "Assessing the Costs and Benefits of Electronic Fetal Monitoring," *Obstet. & Gyn. Survey* 34:627, 1979.
18. Banta, H. D., et al., "Implications of the 1976 Medical Devices Legislation," *Man & Medicine* 3:131, 1978.
19. Banta, H. D., et al., *Toward Rational Technology in Medicine* (New York: Springer Publishing Co., 1981).
20. Barsamian, E. M., "The Rise and Fall of Internal Mammary Artery Ligation in the Treatment of Angina Pectoris and the Lessons Learned," in *Costs, Risks, and Benefits of Surgery,* J. P. Bunker, et al. (eds.) (New York: Oxford University Press, 1977).
21. Bauer, R. A., and Wortzel, L. H., "Doctors Choice: The Physician and His Sources of Information About Drugs," *J. Mktg. Research* 3:40, 1966.
22. Beebe, G. W., "Record Linkage and Needed Improvements in Existing Data Resources," presented at the Banbury Conference on Occupational Cancer, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., Mar. 29-Apr. 2, 1981.
23. Beebe, G. W., "Record Linkage Systems—Canada vs. the United States," *Am. J. Public Health* 70(12):1246, 1980.
24. Beebe, G. W., "Some Aspects of Record Linkage in Epidemiologic Research in the U.S.," paper presented at the Workshop on Computerized Record Linkage in Cancer Epidemiology, Ottawa, Canada, Aug. 8-10, 1979.
25. Behney, C. J., "The Turning Outward of the National Institutes of Health," presented at the 13th Annual Meeting of the Association for the Advancement of Medical Instrumentation, Washington, D.C., Apr. 1, 1978.

26. Berk, A. A., and Chalmers, T. C., "Cost and Efficacy of the Substitution of Ambulatory for Inpatient Care," *N. Eng. J. Med.* 304:393, 1981.
27. Bessant, J., "Innovation Research: To What Extent Can It Provide Answers?" *Planned Innovation* 2(2):52, 1979.
28. Bice, T., et al., *Executive Summary of the Conference on Capital Financing for Health Facilities,* sponsored by the Health Resources Administration and the University of Pittsburgh, Pittsburgh, Pa., Nov. 19-21, 1976.
29. Bilitch, M., Project Director, University of Southern California, Los Angeles, California, *Registry for Implanted Artificial Cardiac Pacemakers: 66 Month Report,* contract No. 223-79-5060, submitted to the Food and Drug Administration, Rockville, Md., Mar. 28, 1980.
30. Birch & Davis Associates, Inc., *IDD Evaluation Project Hospital Field Manual,* contract No. 500-80-0047, submitted to the Health Care Financing Administration, Washington, D.C., 1980.
31. Birch & Davis Associates, Inc., *IDD Grantee Project Synopsis,* contract No. 500-80-0047, submitted to the Health Care Financing Administration, Washington, D.C., Dec. 17, 1980.
32. Birch & Davis Associates, Inc., *Integrated Data Demonstration Evaluation Study,* contract No. 500-80-0047, submitted to the Health Care Financing Administration, Washington, D.C., Oct. 8, 1980.
33. *The Blue Sheet,* "Coordinated Health Information System Feasibility 'Yet To Be Demonstrated,' HHS' Brandt Tells Hill; Cost, Quality, Confidentiality Are Seen As Problems," 24:9, June 10, 1981.
34. Bonchek, L. I., "Are Randomized Trials Appropriate for Evaluating New Operations?" *N. Eng. J. Med.* 301:44, 1979.
35. Breslow, L., University of California, Los Angeles, personal communication, 1981.
36. Bryant, F. B., and Wortman, P. M., "Secondary Analysis: The Case for Data Archives," *Am. Psychol.* 33:381, 1978.
37. Bumpus, F. M., "The Role of Major Medical Clinics in Research," presented at the Project HOPE conference, "The Role of Major Medical Clinics in Meeting Future National Health Needs," Millwood, Va., Mar. 11-13, 1981.
38. Bunker, J. P., Stanford, Calif., personal communication, June 16, 1981.
39. Bunker, J. P., et al., "Assessment of Medical Technology Strategies: Coverage and Reimbursement Options for Experimental Medical and Surgical Procedures," prepared for the Office of Technology Assessment, U.S. Congress, Washington, D.C., September 1981.
40. Bunker, J. P., et al., "Surgical Innovation and Its Evaluation," *Science* 200:937, 1978.
41. Bunning, R. L. "The Delphi Technique: A Projection Tool for Serious Inquiry," in *The 1979 Annual Handbook for Group Facilitators*, J. E. Jones, J. W. Pfeiffer (eds.) (La Jolla, Calif.: University Associates, 1979).
42. Byar, D. P., et al., "Randomized Clinical Trials: Perspectives on Some Recent Ideas," *N. Eng. J. Med.* 295(2):74, 1976.
43. Campbell, D. T., and Boruch, R. F., "Making the Case for Randomized Assignment to Treatments by Considering the Alternatives: Six Ways in Which Quasi-Experimental Evaluations in Compensatory Education Tend To Underestimate Effects," in *Evaluation and Experiment: Some Critical Issues in Assessing Social Programs,* C. A. Bennett and A. A. Lumsdaine (eds.) (New York: Academic Press, 1975).
44. Campbell, D. T., and Erlebacher, A., "How Regression Artifacts in Quasi-Experimental Evaluations Can Mistakenly Make Compensatory Education Look Harmful," in *Compensatory Education: A National Debate,* J. Helmuth (ed.), vol. III of *The Disadvantaged Child* (New York: Brunner/Mazel, 1970).
45. Campbell, D. T., and Stanley, J. C., *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand-McNally, 1966).
46. Caplan, A. L., Hastings Center, Hastings-on-the-Hudson, N.Y., personal communication, November 1981.
47. Caplan, A. L., "Ethical Engineers Need Not Apply: The State of Applied Ethics Today," *Sci. Technol. Human Val.* 6(33):24, 1980.
48. Caplan, A. L., "Kidneys, Ethics, and Politics: Policy Lessons of the ESRD Experience," *J. Health Polit. Policy Law* 6:488, 1981.
49. Caplow, T., *Principles of Organization* (New York: Harcourt, Brace & World, 1964).
50. Center for the Analysis of Health Practices, Harvard School of Public Health, "Impact on Health Costs of NCHCT Recommendations To Withhold Reimbursement for New Medical Procedures," unpublished, February 1981.
51. Chalmers, T. C., "The Clinical Trial," *Milbank Mem. Fund Q.* 59(3):324, 1981.
52. Chalmers, T. C., "Randomization Coronary Artery Surgery," *Ann. Thorac. Surg.* 14:323, 1972.
53. Chalmers, T. C., et al., "Controlled Studies in Clinical Cancer Research," *N. Eng. J. Med.* 287:75, 1972.

54. Chalmers, T. C., et al., "A Method for Assessing the Quality of a Randomized Clinical Trial," *Contr. Clin. Trials* 2:31, 1981.
55. Charles River Associates, Inc. "The Effects of Patent-Term Restoration on the Pharmaceutical Industry," prepared for the Office of Technology Assessment, U. S. Congress, Washington, D.C., May 4, 1981.
56. Charters, A. C., and Bailey, R. N., "Experience With a Simplified Trauma Registry: Profile of Trauma at a University Hospital," *J. Trauma* 19:13, 1979.
57. Childress, J. F., *Priorities in Biomedical Ethics* (Philadelphia: Westminster Press, 1981).
58. Christenson, D. B., and Wertheimer, A. I., "Source of Information and Influence on New Drug Prescribing Among Physicians in an HMO," *Soc. Sci. Med.* 13A:313, 1979.
59. Clarke, R. V. G., and Cornish, D. B., *The Controlled Trial Institutional Research* (London: Home Office Research Studies, 1972).
60. Cochrane, A. L., *Effectiveness and Efficiency: Random Reflections on Health Services* (London: Burgess, 1972).
61. Cole, P., "The Evolving Case-Control Study," *J. Chron. Dis.* 32:15, 1979.
62. Coleman, J. S., et al., *Medical Innovation: A Diffusion Study* (Indianapolis: Bobbs-Merrill, 1966).
63. Committee on the Life Sciences and Social Policy, *Assessing Biomedical Technologies: An Inquiry Into the Nature of the Process* (Washington, D.C.: National Academy of Sciences, 1975).
64. Comroe, J. H., Jr., and Dripps, R. D., *The Top Ten Clinical Advances in Cardiovascular-Pulmonary Medicine and Surgery, 1945-1975,* vol. I, Final Report, DHEW Publication No. (NIH) 78-1521, Jan. 31, 1977.
65. Congressional Research Service, Library of Congress, *Competitive Problems in the Drug Industry: Fixed-Dose Combination Antibiotic Drugs,* prepared for the Select Committee on Small Business, U.S. Senate, May 1, 1979 (Washington, D.C.: U.S. Government Printing Office, 1979).
66. Conner, R. F., "Selecting a Control Group: An Analysis of the Randomization Process in Twelve Social Reform Programs," *Eval. Q.* 1:195, 1977.
67. "Consensus Conference: Diagnosis and Treatment of Reye's Syndrome," *J.A.M.A.* 246:2441, 1981.
68. Cook, T. D., and Campbell, D. T., *Quasi-Experimentation: Design and Analysis of Research in Field Settings* (Chicago: Rand McNally, 1979).
69. Cook, T. D., and Leviton, L. C., "Reviewing the Literature: A Comparison of Traditional Methods With Meta-Analysis," *J. Personality* 48(4):449, 1980.
70. Cook, T. D., and Reichardt, C. S., "Guidelines: Statistical Analysis of Nonequivalent Control Group Designs: A Guide to Some Current Literature," *Evaluation* 3(1-2):136, 1976.
71. Cordray, D. S., and Lipsey, M. W., "Using Patterns of Evidence in Social Program Evaluation: An Overview and Illustration," presented at the meeting of the Evaluation Research Society, Minneapolis, 1979.
72. Coronary Drug Project Research Group, "Influence of Adherence to Treatment and Response of Cholesterol on Mortality in the Coronary Drug Project," *N. Eng. J. Med.* 303:18, 1980.
73. Cowley, M. J., and Block, P. C., "Percutaneous Transluminal Coronary Angioplasty," *Mod. Concepts Cardiovasc. Dis.* 50(5):2, 1981.
74. Crawford, S., "Health Science Libraries in the United States, 1969 to 1979," *J.A.M.A.* 245:237, 1981.
75. Cromwell, J., et al., *Incentives and Decisions Underlying Hospitals' Adoption and Utilization of Major Capital Equipment* (Cambridge, Mass.: Abt Associates, 1975).
76. Cronbach, L. J., and Furby, L., "How Should We Measure 'Change'—Or Should We?" *Psychol. Bull.* 74:68, 1970.
77. Dalkey, N. C., *The Delphi Method: An Experimental Study of Group Opinion* (Santa Monica, Calif.: Rand Corp., 1969).
78. Dalkey, N. C., and Helmer, O., "An Experimental Application of the Delphi Method to the Use of Experts," *Mgt. Sci.* 9(3):458, 1963.
79. D'Arbeloff, D. V., Chairman and Chief Executive Officer, Millipore Corp., statement at Workshop on Innovation and Reimbursement for the Study Strategies for Medical Technology Assessment, Office of Technology Assessment, U.S. Congress, Washington, D.C., May 13, 1981.
80. David, G., Commission on Professional and Hospital Activities, Ann Arbor, Mich., personal communication, May 21, 1981.
81. Deckert, T., et al., "Prognosis of Diabetics With Diabetes Onset Before the Age of Thirty-One," *Diabetologia* 14:363, 1978.
82. Delbecq, A., et al., *Group Techniques for Program Planning* (Glenview, Ill.: Scott, Foresman & Co., 1975).
83. Department of Commerce, National Technical Information Service, *Biomedical Technology and Human Factors Engineering,* Aug. 5, 1980, PB80-909532 (Springfield, Va.: NTIS, 1981).
84. Department of Commerce, National Technical

Information Service, *General Catalog of Information Services, No. 7a* (Springfield, Va.: NTIS, 1981).

85. Department of Commerce, National Technical Information Service, *Health Planning and Health Services Research,* May 19, 1981, PB81-904420 (Springfield, Va.: NTIS, 1981).
86. Department of Commerce, National Technical Information Service, *NTIS Information Services for Productivity, Technology, and Innovation,* information packet (Springfield, Va.: NTIS, 1981).
87. Department of Health, Education, and Welfare, National Center for Health Services Research, *Medical Technology,* publication No. (PHS) 79-3254 (Washington, D.C.: U.S. Government Printing Office, 1979).
88. Department of Health, Education, and Welfare, National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, *The Belmont Report,* DHEW publication No. (OS) 78-0012, 1978.
89. Department of Health, Education, and Welfare, National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, *The Special Study: Implications of Advances in Biomedical and Behavioral Research,* 1978.
90. Department of Health, Education, and Welfare, Office of the Assistant Secretary for Health/Public Health Service, "NCHCT Fact Sheet," n.d.
91. Department of Health, Education, and Welfare, Office of the Assistant Secretary for Planning and Evaluation and the Office of the Assistant Secretary for Health, *Health Technology Management at the Department of Health, Education, and Welfare,* Final Phase I Report for the Secretary, 1977.
92. Department of Health, Education, and Welfare, Public Health Service, Health Resources Administration, Bureau of Health Planning, *The National Health Planning Information Center,* DHEW publication No. (HRA) 79-14007, October 1978.
93. Department of Health, Education, and Welfare, Public Health Service, National Center for Health Services Research, "Clinical Databanks—An Epidemiological Approach to Technology Assessment and Consensus Development," presented at the XII International Conference on Medical and Biological Engineering, Jerusalem, Israel, Aug. 19-24, 1979.
94. Department of Health, Education, and Welfare, Public Health Service, National Center for Health Services Research, *Computer Applications in Health Care,* DHEW publication No. (PHS) 79-3251, June 1979.
95. Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health, "Consensus Development Conference Statement," *N. Eng. J. Med.* 304:680, 1981.
96. Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health, *Consensus Development Conference Summaries,* vol. 1, 1977-1978.
97. Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health, *Consensus Development Conference Summaries,* vol. 2, 1979.
98. Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health, *Consensus Development Conference Summaries,* vol. 3, 1980.
99. Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health, *NIH Inventory of Clinical Trials: Fiscal Year 1975,* vols. I and II, 1977.
100. Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health, "Program Description and Data Format: Cancer Surveillance Epidemiology and End Results (SEER) Reporting Program," unpublished, April 1976.
101. Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health, "The Responsibilities of NIH at the Health Research/Health Care Interface," draft, Feb. 14, 1978.
102. Department of Health and Human Services, Health Care Financing Administration, "Health Care Information System of the 80's: HCFA's Perspective and Initiatives," discussion paper, May 14, 1980.
103. Department of Health and Human Services, Health Care Financing Administration, "Operating Guidelines for Determining Medical Necessity of Medicare Items and Services," draft, Feb. 2, 1971.
104. Department of Health and Human Services, Health Care Financing Administration, Office of Research, Demonstrations, and Statistics, *Grants for Research and Demonstrations, Fiscal Year 1981,* HCFA Publication No. 03074, December 1980.
105. Department of Health and Human Services, Health Care Financing Administration, Office of Research, Demonstrations, and Statistics, "A Guide to ORDS Publications," mimeo, August 1980.

106. Department of Health and Human Services, Health Care Financing Administration, Office of Research, Demonstrations, and Statistics, "ORDS Research Publications," mimeo, n.d.
107. Department of Health and Human Services, Health Care Financing Administration, Office of Research, Demonstrations, and Statistics, "Research Publications Program," mimeo, n.d.
108. Department of Health and Human Services, Office of the Assistant Secretary for Health/Public Health Service, *Maternal Serum Alpha-Fetoprotein: Issues in the Prenatal Screening and Diagnosis of Neural Tube Defects,* National Center for Health Care Technology Conference Proceedings, Washington, D.C., July 28-30, 1980.
109. Department of Health and Human Services, Office of the Assistant Secretary for Health/Public Health Service, National Center for Health Care Technology, "Hemodialysis for Treatment of Schizophrenia," memorandum to the Health Care Financing Administration, Sept. 18, 1980.
110. Department of Health and Human Services, Office of Assistant Secretary for Health/Public Health Service, National Center for Health Care Technology, "Procedures, Priorities and Policy for Assessment of Health Care Technology," Sept. 1, 1980.
111. Department of Health and Human Services, Office of the Assistant Secretary for Health/Public Health Service, National Center for Health Care Technology, "The Process for Assessing Technologies," Jan. 21, 1981.
112. Department of Health and Human Services, Office of the Assistant Secretary for Health/Public Health Service, National Center for Health Care Technology, *Report to the Congress,* 1980.
113. Department of Health and Human Services, Office of the Assistant Secretary for Health/Public Health Service, National Center for Health Services Research, *Program Description,* Jan. 9, 1981.
114. Department of Health and Human Services, Office of the Assistant Secretary for Health/Public Health Service, National Center for Health Statistics, *Data Systems of the National Center for Health Statistics,* DHEW publication No. (PHS) 80-1247, 1980.
115. Department of Health and Human Services, Office of the Assistant Secretary for Health/Public Health Service, National Center for Health Statistics, *Directions for the '80s: Final Report of the Panel To Evaluate the Cooperative Health Statistics System,* DHHS publication No. (PHS) 80-1204, August 1980.
116. Department of Health and Human Services, Office of the Assistant Secretary for Health/Public Health Service, National Center for Health Statistics, *Environmental Health: A Plan for Collecting and Coordinating Statistical and Epidemiologic Data,* DHHS publication No. (PHS) 80-1248, 1980.
117. Department of Health and Human Services, Office of the Assistant Secretary for Health/Public Health Service, National Center for Health Statistics, *Health: United States, 1980,* DHHS publication No. (PHS) 81-1232, October 1980.
118. Department of Health and Human Services, Office of the Assistant Secretary for Health/Public Health Service, National Council for Health Care Technology, *Annual Report, Fiscal Year 1979,* n.d.
119. Department of Health and Human Services, Office of the Assistant Secretary for Health/Public Health Service, National Council for Health Care Technology, minutes of meeting, Washington, D.C., Aug. 13, 1981.
120. Department of Health and Human Services, Office of the Assistant Secretary for Health/Public Health Service, Technology Coordinating Committee, minutes of meeting, Apr. 27, 1981.
121. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation, "Project Share: A National Clearinghouse for Improving the Management of Human Services," DHHS publication No. OS-76-221, 1981.
122. Department of Health and Human Services, Public Health Service, Food and Drug Administration, Office of Planning and Evaluation, Economics Staff, "Baseline Data on Medical Device Industries in the Census of Manufacturers," OPE study 53, 1980.
123. Department of Health and Human Services, Public Health Service, Health Resources Administration, Bureau of Health Planning, "The National Health Planning Information Center: Its Conception, Development, and Organization," mimeo, n.d.
124. Department of Health and Human Services, Public Health Service, Health Resources Administration, Bureau of Health Planning, "Services Provided by the National Health Planning Information Center," mimeo, June 1980.
125. Department of Health and Human Services, Public Health Service, National Institutes of Health, *NIH Public Advisory Groups: Authority, Structure, Functions, Members,* publication No. (NIH) 79-11, Jan. 1, 1979.

126. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Heart, Lung, and Blood Institute, Office of Information, "Coronary Artery Surgery Study," unpublished, 1981.
127. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Heart, Lung and Blood Institute, Office of Information, "The Framingham Heart Study," unpublished, 1981.
128. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Heart, Lung, and Blood Institute, Office of Information, "Multiple Risk Factor Intervention Trial for the Prevention of Coronary Heart Disease," n.d.
129. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Library of Medicine, Lister Hill National Center for Biomedical Communications, "The Hepatitis Knowledge Base," *Ann. Int. Med.* 93(1)(Part 2):165, 1980.
130. Dowling, H. F., cited in *Competitive Problems in the Drug Industry: Fixed-Dose Combination Antibiotic Drugs*, prepared by the Congressional Research Service, Library of Congress, for the Select Committee on Small Business, U.S. Senate, May 1, 1979 (Washington, D.C.: U.S. Government Printing Office, 1979).
131. DuMouchel, W. H., and Harris, J. E., "Bayes Methods for Combining Cancer Experiments in Man and Other Species," *J. Amer. Statist. Assoc.*, in press, 1982.
132. Duncan, K. A., Palo Alto, Calif. "The Trend Toward a National Health Information System in the United States," unpublished paper, n.d.
133. "Editorial: New Insulin-Delivery Systems for Diabetics," *Lancet* 1:1275, 1979.
134. Einhorn, H. J., and Hogarth, R. M., "Confidence in Judgment: Persistence in the Illusion of Validity," *Psychol. Rev.* 85:395, 1978.
135. Enzer, S., et al., *Some Societal Changes by 1985 and Their Impacts on Time/Money Budget* (Middletown, Conn.: Institute for the Future, 1971).
136. European Coronary Surgery Study Group, "Coronary-Artery Bypass Surgery in Stable Angina Pectoris: Survival at Two Years," *Lancet* 1:889, 1979.
137. Eysenck, H. F., "An Exercise in Mega-Silliness," *Am. Psychol.* 33:517, 1978.
138. Federal Coordinating Committee on Science, Engineering and Technology, Ad Hoc Committee on Scientific and Technical Information, "Report of the Working Group on Private Sector/Government Relations for Scientific and Technical Information," mimeo obtained from the National Science Foundation, Washington, D.C., June 1980.
139. Feer, H., et al., *Compr. Psychiatry* 1:338, 1960, cited in S. C. Schulz, et al., "Dialysis in Schizophrenia: A Double-Blind Evaluation," *Science* 211:1066, 1981.
140. Fein, R., "On Measuring Economic Benefits of Health Programs," in *Ethics and Health Policy*, R. M. Veatch and R. Branson (eds.) (Cambridge, Mass.: Ballinger, 1976).
141. Feinstein, A. P., "A Survey of the Statistical Procedures in General Medical Journals," *Clin. Pharmacol. Therapeut.* 15:97, 1975.
142. Feldstein, M., and Taylor, A., *The Rapid Rise of Hospital Costs* (Washington, D.C.: President's Council on Wage and Price Stability, 1977).
143. Fineberg, H. V., "Gastric Freezing—A Study of Diffusion of Medical Innovation," in *Medical Technology and the Health Care System* (Washington, D.C.: National Academy of Sciences, 1979).
144. Fineberg, H. V., and Hiatt, H. H., "Evaluation of Medical Practices: The Case for Technology Assessment," *N. Eng. J. Med.* 301:1086, 1970.
145. Fineberg, H. V., et al., "Acquisition and Application of New Medical Knowledge by Anesthesiologists: Three Recent Examples," *Anesthesiology* 48:430, 1978.
146. Fishbein, H. A., State of Rhode Island Department of Health, Providence, R.I., personal communication, May 7, 1981.
147. Fisher, B., "Breast-Cancer Management: Alternatives to Radical Mastectomy," *N. Eng. J. Med.* 301:326, 1979.
148. Fletcher, R. H., and Fletcher, S. W., "Clinical Research in General Medical Journals," *N. Eng. J. Med.* 301:180, 1979.
149. Frank, J. D., *Persuasion and Healing* (Baltimore: Johns Hopkins University Press, 1973).
150. Frankina, W. K., *Ethics* (Englewood Cliffs, N.J.: Prentice-Hall, 1973).
151. Freeman, C., *The Role of Small Firms in Innovation in the United Kingdom Since 1945*, report to the Bolton Committee of Inquiry on Small Firms, research report No. 6 (London: Her Majesty's Stationery Office, 1971).
152. Freeman, G., Blue Cross/Blue Shield Association, Chicago, Ill., personal communication, May 1981.
153. Freiman, J. A., et al., "The Importance of Beta, the Type II Error and Sample Size in the Design and Interpretation of the Randomized Control Trial," *N. Eng. J. Med.* 299:690, 1978.

154. General Accounting Office, U.S. Congress, *FDA Drug Approval: A Lengthy Process That Delays the Availability of Important New Drugs,* report by the Comptroller General of the U.S. to the Subcommittee on Science, Research, and Technology, House Committee on Science and Technology, U.S. Congress, May 28, 1980.

155. General Accounting Office, U.S. Congress, *OPM Should Promote Medical Necessity Programs for Federal Employees' Health Insurance* (Washington, D.C.: U.S. Government Printing Office, July 29, 1980).

156. Gerstenfeld, A., "A Study of Successful Projects, Unsuccessful Projects, and Projects in Process in West Germany," *IEEE Trans. Eng. Mgt.* EM-23(3):116, 1976.

157. Gifford, R. H., and Feinstein, A. R., "A Critique of Methodology in Studies of Anticoagulant Therapy for Acute Myocardial Infarction," *N. Eng. J. Med.* 280:351, 1969.

158. Gilbert, J. P., et al., "Assessing Social Innovations: An Empirical Base for Policy, in *Evaluation and Experiment: Some Critical Issues in Assessing Social Programs,* C. A. Bennett and A. A. Lumsdaine (eds.) (New York: Academic Press, 1975).

159. Gilbert, J. P., et al., "Progress in Surgery and Anesthesia: Benefits and Risks of Innovative Therapy," in *Costs, Risks, and Benefits of Surgery,* J. P. Bunker, et al. (eds.) (New York: Oxford University Press, 1977).

160. Gilbert, J. P., et al., "Statistics and Ethics in Surgery and Anethesia," *Science* 198:684, 1977.

161. Gillings, D., et al., "Analysis of Interrupted Time Series Mortality Trends: An Example To Evaluate Regionalized Perinatal Care," *Am. J. Public Health* 71:38, 1981.

162. Ginsburg, P., "Resource Allocation in the Hospital Industry: The Role of Capital Financing," *Soc. Sec. Bull.* 35:20, 1972.

163. Glaser, E. M., "Using Behavioral Science Strategies for Defining the State of the Art," *J. Applied Behav. Sci.* 16:79, 1980.

164. Glass, G. V., "Integrating Findings: The Meta-Analysis of Research," in *Review of Research in Education,* vol. 5, L. S. Shulman (ed.) (Itasca, Ill.: Peacock, 1977).

165. Glass, G. V., "Primary Secondary, and Meta-Analysis of Research," *Educ. Res.* 5:3, 1976.

166. Glass, G. V., and Smith, M. L., "Meta-Analysis of Research on Class Size and Achievement," *Educ. Evaluat. Policy Analysis* 1:2, 1979.

167. Glass, G. V., et al., *The Design and Analysis of Time-Series Experiments* (Boulder, Colo.: Colorado Associated University Press, 1975).

168. Goldschmidt, P., Veterans Administration, Washington, D.C., personal communication, January 1982.

169. Gordis, L., et al., "Pitfalls in Evaluating the Impact of Coronary Care Units on Mortality From Myocardial Infarctions," *Johns Hopkins Med. J.* 141:287, 1977.

170. Gore, S. M., "Assessing Clinical Trials—First Steps," *Br. Med. J.* 282:1605, 1981.

171. Gore, S. M., "Assessing Clinical Trials—Trial Size," *Br. Med. J.* 282:1687, 1981.

172. Gore, S. M., et al., "Misuse of Statistical Methods: Critical Assessment of Articles in BMJ From January to March 1976," *Br. Med. J.* 1:85, 1977.

173. Gottman, J. M., "N-of-One and N-of-Two Research in Psychotherapy," *Psychol. Buil.* 80:93, 1973.

174. Grabowski, H. G., et al., "The Effects of Regulatory Policy on the Incentives To Innovate: An International Comparative Analysis," in *Impact of Public Policy on Drug Innovation and Pricing,* S. A. Mitchell and E. A. Links (eds.) (Washington, D.C.: The American University, 1976).

175. Grace, N. D., et al., "The Present Status of Shunts for Protal Hypertension Incirrhosis," *Gastroent.* 50:684, 1966.

176. *The Gray Sheet* "Restricted Device Case-by-Case Alternative Unanimously Urged by Associations," 7(4):I&W8, Jan. 26, 1981.

177. *The Gray Sheet,* "Mandatory Experience Reporting Proposal Unanimously Criticized by Manufacturers," 7(6), Feb. 9, 1981.

178. Greer, A. L., "Advances in the Study of Diffusion of Innovation in Health Care Organizations," *Milbank Mem. Fund Q.* 55:505, 1977.

179. Hale, W. E., et al., "Drug Use in a Geriatric Population," *J. Am. Geriat. Soc.* 23(8):374, 1979.

180. Hammermeister, K. E., et al., "Effect of Aortocoronary Saphenous Vein Bypass Grafting on Death and Sudden Death," *Am. J. Cardiol.* 39:925, 1977.

181. Hammermeister, K. E., et al., "Evidence From a Nonrandomized Study That Coronary Surgery Prolongs Survival in Patients With Two-Vessel Coronary Disease," *Circulation* 59:430, 1979.

182. Harris, J. E., *Macro-Experiments Versus Micro-Experiments for Health Policy* (No. 286A) (Cambridge, Mass.: Massachusetts Institute of Technology Press, 1981).

183. Hayes, R. H., and Abernathy, W. J., "Managing Our Way to Economic Decline," *Harvard Bus. Rev.* 58(4):68, 1980.
184. Hecht, A., "Drug Effects on the Eyes," *FDA Consumer,* November 1978, p. 14.
185. Hecht, A., "Tracking the Skin's Reactions to Drugs," *FDA Consumer,* April 1981, p. 5.
186. Hedges, L. V., and Olkin, I., "Vote-Counting Methods in Research Synthesis," *Psychol. Bull.* 88:359, 1980.
187. Hill, A. B., *Controlled Clinical Trials* (Springfield, Ill.: Charles C. Thomas, 1960).
188. Hill, J. D., et al., "A Randomised Trial of Home-Versus-Hospital Management for Patients With Suspected Myocardial Infarction," *Lancet* 1:837, 1978.
189. Hobbins, J. C., et al., "The Fetal Monitoring Debate," *Pediatrics* 63:942, 1979.
190. Hoehler, F. K., et al., "Spinal Manipulation for Low Back Pain," *J.A.M.A.* 245:1835, 1981.
191. Horwitz, R. I., and Feinstein, A. R., "Alternative Analytic Methods for Case-Control Studies of Estrogens and Endometrial Cancer," *N. Eng. J. Med.* 299(20):1089, 16, 1978.)
192. Horwitz, R. I., et al., "Necropsy Diagnosis of Endometrial Cancer and Detection—Bias in Case/Control Studies," *Lancet* 2:66, 1981.
193. Hubbard, W. N., Jr., President, Upjohn Co., testimony at the hearings on *The Depo-Provera Debate* before the Select Committee on Population, U.S. House of Representatives, Aug. 8, 1978 (Washington, D.C.: U.S. Government Printing Office, 1978).
194. Hulka, B. S., et al., "Physician Management in Primary Care," *Am. J. Public Health* 66:1173, 1966.
195. Hummel-Rossi, B., and Weinberg, S. L., "Practical Guidelines in Applying Current Theories to the Measurement of Change," *JSAS Catalog of Selected Documents in Psychology* (Ms. No. 916) 5:226, 1975.
196. Hutchinson, G. B., and Rothman, K. J., "Correcting a Bias?" *N. Eng. J. Med.* 299:1129, 1978.
197. Institute of Medicine, National Academy of Sciences, *Conference Summary: Evaluating Medical Technologies in Clinical Use* (Washington, D.C.: NAS, 1981).
198. Institute of Medicine, National Academy of Sciences, *Reliability of Hospital Discharge Abstracts* (Washington, D.C.: NAS, 1980).
199. Institute of Medicine, National Academy of Sciences, *Reliability of Medicare Hospital Discharge Records* (Washington, D.C.: NAS, November 1977).
200. Jick, H., et al., "Vaginal Spermicides and Congenital Disorders," *J.A.M.A.* 245(13):1329, 1981.
201. Joint Commission on Prescription Drug Use, Inc., *Report of the Joint Commission on Prescription Drug Use: Appendix II, Minutes of Meetings,* minutes of June 20-21, 1978 (Rockville, Md.: Jan. 23, 1980).
202. Jones, R. R., et al., "Time-Series Analysis in Operant Research," *J. Applied Behav. Analysis* 10:151, 1977.
203. Jones, S. E., "Adjuvant Chemotherapy for Breast Cancer," *J.A.M.A.* 245:1527, 1981.
204. Jonsen, A. R., "The Totally Implantable Artificial Heart," *Hast. Cent. Rpt.* 3(5):1, 1973.
205. Jonsen, A. R., and Butler, L. H., "Public Ethics and Policy Making," *Hast. Cent. Rpt.* 5(4):19, 1975.
206. Joreskog, K. G., "A General Approach to Confirmatory Maximum Likelihood Factor Analysis," *Psychometrika* 34:183, 1969.
207. Kaiser Foundation Health Plan, Oakland, Calif., "Kaiser-Permanente Medical Care Program Annual Report 1980," n.d.
208. Katz, E., et al., "Traditions of Research on the Diffusion of Innovation," *Amer. Sociol.* 28:237, 1963.
209. Kennedy, R. H., et al., "Neurologic Computed Tomography in a Defined Population Group," *Radiology* 106:1, 1979.
210. Kennedy, R. H., et al., "Trends in Cardiac Surgery: Five-Year Study of a Defined Population," *Thoracic Cardiovasc. Surg.* 80(5):702, 1980.
211. Kennedy, R. H., et al., "Use of the Cardiac-Catheterization Laboratory in a Defined Population," *N. Eng. J. Med.* 303:1273, 1980.
212. Kimberly, J. R., "Hospital Adoption of Innovation: The Role of Integration Into External Information Environments," *J. Health Soc. Behav.* 19:361, 1978.
213. Kimberly, J. R., "Managerial Innovation," in *Handbook of Organizational Design,* P. C. Nystrom and W. H. Starbuck (eds.) (New York: Oxford University Press, 1981).
214. Kohrman, A., Washington, D.C., personal communication, 1981.
215. Kolata, G. B., "Dilemma in Cancer Treatment," *Science* 209:792, 1980.
216. Krol, R. A., "A Meta-Analysis of Comparative Research on the Effects of Desegregation on Academic Achievement," unpublished doctoral dis-

sertation, University of Michigan, Ann Arbor, Mich., 1979.

217. Kuhn, T. S., *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1970).
218. Kunitz, S., et al., "Pilot Data Bank Networks for Neurological Disorders," in *Third Annual Proceedings of the Symposium of Medical Care* (Long Beach, Calif.: IEEE Computer Society, 1979).
219. Langrish, J., et al., *Wealth From Knowledge* (London: Macmillan, 1972).
220. Lave, L. B., and Seskin, E. P., "Epidemiology, Causality and Public Policy," *Am. Sci.* 67:178, 1979.
221. Lebacqz, K., "Bioethics: Some Challenges From a Libertarian Perspective," presented at the Conference on Faith, Science, and the Future, Cambridge, Mass., July 1979.
222. Levy, R. I., and Sondik, E. J., "Decision-Making in Planning Large-Scale Comparative Studies," *Ann. N. Y. Acad. Sci.* 304:441, 1978.
223. Levy, R. I., and Sondik, E. J., "The Management of Biomedical Research: A Case in Point," in *Biomedical Innovation,* E. B. Roberts, et al. (eds.) (Cambridge, Mass.: MIT Press, 1981).
224. Levy, D., New York Hospital, personal communication, April 1981.
225. Levy, D., et al., "Prognosis in Nontraumatic Coma," *Ann. Int. Med.* 94(3):293, 1981.
226. Light, R. J., and Smith, P. V., "Accumulating Evidence: Procedures for Resolving Contradictions Among Research Studies," *Harvard Educ. Rev.* 41:429, 1976.
227. Lohr, K. N., et al., *Peer Review and Technology Assessment in Medicine,* prepared for the Office of Technology Assessment, U.S. Congress (Santa Monica, Calif.: Rand Corp., August 1981).
228. Lowe, C. U., Office of Medical Applications of Research, National Institutes of Health, Bethesda, Md., personal communication, May 15, 1981.
229. Lowe, C. U., "The Consensus Development Programme: Technology Assessment at the National Institutes of Health," *Br. Med. J.* 28:1583, 1980.
230. Luborsky, L., et al., "Comparative Studies of Psychotherapies: Is It True That Everyone Has Won and All Must Have Prizes?" *Arch. Gen. Psychiatry* 32:995, 1975.
231. Malkiel, B. G., "Productivity: The Problem Behind the Headlines," *Harvard Bus. Rev.* 57(3):81, 1980.
232. Malone, W. F., et al., "The Delphi Technique as an Instrument in Dental Curriculum Revision," *J. Dent. Educ.* 40:298, 1976.
233. Manning, P. R., and Denson, T. A., "How Cardiologists Learn About Echocardiography: A Reminder for Medical Educators and Legislators," *Ann. Int. Med.* 91:469, 1979.
234. Manning, P. R., and Denson, T. A., "How Internists Learned About Cimetidine," *Ann. Int. Med.* 92:690, 1980.
235. Mansfield, E., et al., *The Production and Application of New Industrial Technology* (New York: Norton, 1977).
236. Mather, H. G., et al., "Myocardial Infarction: A Comparison Between Home and Hospital Care for Patients," *Br. Med. J.* 1:925, 1976.
237. McCleary, R., and Hay, R. A., Jr., *Applied Time Series for the Social Sciences* (Beverly Hills, Calif.: Sage Publications, 1980).
238. McGinnis, W., Department of Defense, Tri-Service Medical Information Systems Program Office, Bethesda, Md., personal communication, Dec. 10, 1981.
239. McGuire, W. J., "The Nature of Attitudes and Attitude Change," in *Handbook of Social Psychology,* vol. 3., G. Lindzey and E. Aronson (eds.) (Reading, Mass.: Addison-Wesley, 1969).
240. McNeer, J. F., et al., "The Nature of Treatment Selection in Coronary Artery Disease: Experience With Medical and Surgical Treatment of a Chronic Disease," *Circulation* 49:606, 1974.
241. McNeil, G. J., and Adelstein, S. J., "Determining the Value of Diagnostic and Screening Tests," *J. Nucl. Med.* 17:439, 1976.
242. Meier, P., University of Chicago, personal communication, Dec. 4, 1980.
243. Meier, P., "The Biggest Public Health Experiment Ever: The 1954 Field Trial of the Salk Polimyelitis Vaccine," in *Statistics: A Guide to the Unknown,* J. M. Tanur, et al. (eds.) (San Francisco: Holden-Day, 1972).
244. Merrill, R. A., "FDA Modes of Regulation and Their Relationship to Private Sector Innovation," prepared for the Office of Technology Assessment, U.S. Congress, Washington, D.C., 1980.
245. Miao, L. L. "Gastric Freezing: An Example of the Evaluation of Medical Therapy by Randomized Clinical Trials," in *Costs, Risks, and Benefits of Surgery,* J. P. Bunker, et al. (eds.) (New York: Oxford University Press, 1977).
246. Milholland, A. V., et al., "Medical Assessment by a Delphi Group Opinion Technique," *N. Eng. J. Med.* 288:1272, 1973.
247. Miller, M. J., "FDA Voluntary Standards Endorsement Policy: Victim of Legal Homicide," *Med. Instrumentation* 14:331, 1980.
248. Mitchell, A. A., et al., "Adverse Drug Effects and

Drug Surveillance," in *Pediatric Pharmacology—Therapeutic Principles in Practice,* S. J. Yaffe (ed.) (New York: Grune & Stratton, 1980).

249. Moloney, T. W., and Rogers, D. E., "Medical Technology: A Different View of the Contentious Debate Over Costs," *N. Eng. J. Med.* 301(26): 1413, 1979.
250. Moscovice, I., et al., "Health Services Research for Decision-Makers: The Use of the Delphi Technique To Determine Health Priorities," *J. Health Polit. Policy Law* 3:388, 1977.
251. Mosenkis, R., Emergency Care Research Institute, Philadelphia, Pa., personal communication, August 1981.
252. Mosteller, F., "Innovation and Evaluation," *Science* 211(4485):881, 1981.
253. Muller, J. E., et al., "Let's Not Let the Genie Escape From the Bottle—Again," *N. Eng. J. Med.* 304:1294, 1981.
254. Murphy, M. L., et al., "Special Correspondence: A Debate on Coronary Bypass," *N. Eng. J. Med.* 297:1470, 1977.
255. Murphy, M. L., et al., "Treatment of Chronic Stable Angina: A Preliminary Report of Survival Data of the Randomized Veterans Administration Cooperative Study," *N. Eng. J. Med.* 297:621, 1977.
256. Myers, S., and Marquis, D. G., *Successful Industrial Innovations,* NSF 69-71 (Washington, D.C.: National Science Foundation, 1969).
257. National Science Foundation, *Science Indicators 1978* (Washington, D.C.: NSF, 1978).
258. Nelson, N. M., et al., "A Randomized Clinical Trial of the Leboyer Approach to Childbirth," *N. Eng. J. Med.* 302:655, 1980.
259. Neustadt, R. D., and Fineberg, H. V., *The Swine Flu Affair: Decision-Making on a Slippery Disease,* GPO stock No. 017-000-002104 (Washington, D.C.: U.S. Government Printing Office, 1978).
260. "New Device Introductions on the Rise," *Dev. & Diag. Letter* 7:3, Aug. 12, 1980.
261. Nino, H., Food and Drug Administration, Bureau of Medical Devices, Rockville, Md., personal communication, January 1981.
262. Nisbett, R. E., and Ross, L. D., *Human Inference: Strategies and Shortcomings of Social Judgment* (Englewood Cliffs, N.J.: Prentice-Hall, 1980).
263. Nobrega, F. T., et al., Mayo Clinic, Rochester, Minn., "Hospital Use in a Circumscribed Community," unpublished paper, 1980.
264. The Norwegian Multicenter Study Group, "Timolol-Induced Reduction in Mortality and Reinfarction in Patients Surviving Acute Myocardial Infarction," *N. Eng. J. Med.* 304:801, 1981.
265. Office of Technology Assessment, U.S. Congress, *Annual Report to the Congress for 1977* (Washington, D.C.: U.S. Government Printing Office, 1978).
266. Office of Technology Assessment, U.S. Congress, *Assessing the Efficacy and Safety of Medical Technologies,* GPO stock No. 052-003-00593-0 (Washington, D.C.: U.S. Government Printing Office, September 1978).
267. Office of Technology Assessment, U.S. Congress, *Assessment of Technologies for Determining Cancer Risks From the Environment,* GPO stock No. 052-003-00832-7 (Washington, D.C.: U.S. Goverment Printing Office, June 1981).
268. Office of Technology Assessment, U.S. Congress, *Computer Technology in Medical Education and Assessment,* GPO stock No. 052-003-00704-5, (Washington, D.C.: U.S. Government Printing Office, September 1979).
269. Office of Technology Assessment, U.S. Congress, *Development of Medical Technology: Opportunities for Assessment,* GPO stock No. 052-003-00217-5 (Washington, D.C.: U.S. Government Printing Office, August 1976).
270. Office of Technology Assessment, U.S. Congress, *The Implications of Cost-Effectiveness Analysis of Medical Technology,* GPO stock No. 052-003-00765-7 (Washington, D.C.: U.S. Government Printing Office, August 1980).
271. Office of Technology Assessment, U.S. Congress, *The Implications of Cost-Effectiveness Analysis of Medical Technology/Background Paper #1: Methodological Issues and Literature Review,* GPO stock No. 052-003-00780-1 (Washington, D.C.: U.S. Government Printing Office, September 1980).
272. Office of Technology Assessment, U.S. Congress, *The Implications of Cost-Effectiveness Analysis of Medical Technology/Background Paper #2: Case Studies of Medical Technologies* (Washington, D.C.: U.S. Government Printing Office, 1981).
273. Office of Technology Assessment, U.S. Congress, *The Implications of Cost-Effectiveness Analysis of Medical Technology/Background Paper #3: The Efficacy and Cost Effectiveness of Psychotherapy,* GPO stock No. 052-003-00780-5 (Washington, D.C.: U.S. Government Printing Office, October 1980).
274. Office of Technology Assessment, U.S. Congress, *The Implications of Cost-Effectiveness Analysis*

of Medical Technology/Background Paper #4: The Management of Health Care Technology in Ten Countries, GPO stock No. 052-003-00780-3 (Washington, D.C.: U.S. Government Printing Office, October 1980).

275. Office of Technology Assessment, U.S. Congress, *Medical Technology Under Proposals To Increase Competition in Health Care,* in press, 1982.
276. Office of Technology Assessment U.S. Congress, *MEDLARS and Health Information Policy,* in press, 1982.
277. Office of Technology Assessment, U.S. Congress, "The National Center for Health Care Technology and Medical Technology Assessment Policy," staff paper, Mar. 16, 1981.
278. Office of Technology Assessment, U.S. Congress, "The National Library of Medicine," staff paper, Mar. 17, 1981.
279. Office of Technology Assessment, U.S. Congress, *Policy Implications of the Computed Tomography (CT) Scanner,* GPO stock No. 052-003-00565-4, (Washington, D.C.: U.S. Government Printing Office, August 1978).
280. Office of Technology Assessment, U.S. Congress, *Policy Implications of the Computed Tomography (CT) Scanner: An Update,* GPO stock No. 052-003-00793-2, (Washington, D.C.: U.S. Government Printing Office, January 1981).
281. Office of Technology Assessent, U.S. Congress, *Postmarketing Surveillance of Prescription Drugs,* in press, 1982.
282. Office of Technology Assessment, U.S. Congress, *Selected Topics in Federal Health Statistics,* GPO stock No. 052-003-00683-9 (Washington, D.C.: U.S. Government Printing Office, June 1979).
283. Office of Technology Assessment, U.S. Congress, *Technological Innovation and Health, Safety, and Environmental Regulations,* in press, 1982.
284. Okie, S., "Hard To Put Price on Heart Surgery's Value," *The Washington Post,* Apr. 27, 1981, A13.
285. Palmour, R. M., and Ervin, F. R., "Biochemical and Behavioral Studies of a Peptide Derived From the Dialysate of Psychiatric Patients," presented at the 24th Annual Meeting of the American Society for Artificial Internal Organs, April 1978.
286. Patton, M. Q., *Qualitative Evaluation Methods* (Beverly Hills, Calif.: Sage Publications, 1979).
287. Perry, S., and Kalberer, J. T., Jr., "The NIH Consensus-Development Program and the Assessment of Health Technologies: The First Two Years," *N. Eng. J. Med.* 303:169, 1980.
288. Peterson, R. D., and MacPhee, C. R., *Economic Organization in Medical Equipment and Supply* (Lexington, Mass.: D. C. Heath & Co., 1973).
289. Peto, R., et al., "Design and Analysis of Randomized Clinical Trials Requiring Prolonged Observation of Each Patient: 1. Introduction and Design," *Br. J. Cancer* 34:585, 1976.
290. Pickup, J. C., and Keen, H., "Continuous Subcutaneous Insulin Infusion: A Developing Tool in Diabetes Research," *Diabetologia* 18:1, 1980.
291. Pickup, J. C., et al., "Continuous Subcutaneous Insulin Infusion: Improved Blood-Glucose and Intermediary-Metabolite Control in Diabetics," *Lancet* 1:1255, 1979.
292. Pillemer, D. B., and Light, R. J., "Synthesizing Outcomes: How To Use Research Evidence From Many Studies," *Harvard Educ. Rev.* 50:176, 1980.
293. Policy Research, Inc., Baltimore, Md., *Medical Practice Information Demonstration Project,* final report to the Office of the Assistant Secretary for Health, Department of Health, Education, and Welfare, contract No. 282-77-0068 GS, December 1979.
294. Posavac, E. J., "Evaluations of Patient Education Programs: A Meta-Analysis," *Eval. Health Profess.* 3:47, 1980.
295. Powledge, T. M., and Fletcher, J., "Guidelines for the Ethical, Social and Legal Issues in Prenatal Diagnosis," *N. Eng. J. Med.* 300:168, 1979.
296. President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research, *Defining Death: Medical, Legal and Ethical Issues in the Determination of Death* (Washington, D.C.: U.S. Government Printing Office, July 1981).
297. Preston, T. A., *Coronary Artery Surgery: A Critical Review* (New York: Raven, 1977).
298. Quinn, J. C., "Why Executives Think Short," *Newsweek,* July 13, 1981.
299. Radzius, J. R., "FDA Regulation of Drugs, Devices," *AORN J.* 24:254, 1976.
300. Ramroth, D. N. (ed.), *Proceedings From the Medical and Scientific Section Meeting on Getting New Products to the Market* (Washington, D.C.: Health Industry Manufacturers Association, 1979).
301. Rapoport, J., "Diffusion of Technological Innovation Among Non-Profit Firms: A Case Study of Radioisotopes in U.S. Hospitals," *J. Econ. Bus.* 301:108, 1978.
302. Registry for Research on Hormonal Transplacen-

tal Carcinogenesis, *1978 Newsletter,* Chicago, n.d.

303. Registry for Research on Hormonal Transplacental Carcinogenesis, *1979-80 Newsletter,* Chicago, n.d.
304. Reich, W. T. (ed.), *Encyclopedia of Bioethics* (New York: Free Press, 1978).
305. Reiss, J. B., "Issues in the Cost and Regulation of New Medical Technologies and Procedures: Heart Transplants as a Case Study," presented to the Harvard-MIT Conference on Medical Technology, Boston, Apr. 13-16, 1980.
306. Reiss, J. B., "Technology Assessment and Paying for Medical Services Provided Patients," prepared for the Office of Technology Assessment, U.S. Congress, Washington, D.C., April 1981.
307. Relman, A. S., "Assessment of Medical Practices," *N. Eng. J. Med.* 303(3):153, 1980.
308. Rennie, R. A., "Consensus Statements," *N. Eng. J. Med.* 303:153, 1980.
309. Research News, *The Diabetes Research and Training Center: New Assault on an Old Enemy* (Ann Arbor: University of Michigan, May, 1979).
310. Rettig, R. A., "The Federal Government and Medical Technology: Crossing Policy and Management Thresholds," *Policy Sci.* 11:343, 1980.
311. Riecken, H. W., and Boruch, R. F. (eds.), *Social Experimentation: A Method for Planning and Evaluating Social Intervention* (New York: Academic Press, 1975).
312. Rindskopf, D., "Using Structural Equation Models To Analyze Nonexperimental Data," in *Reanalyzing Program Evaluations: Policies and Practices for Secondary Analysis of Social and Educational Programs,* R. F. Boruch, et al. (eds.) (San Francisco: Jossey-Bass, in press).
313. Roberts, E. B., "Influences on Innovation: Extrapolations to Biomedical Technology," in *Biomedical Innovation,* E. B. Roberts, et al. (eds.) (Cambridge, Mass.: MIT Press, 1981).
314. Robinson, J. A., "Coronary Care Unit Versus Hospital Mortality in Acute Myocardial Infarction," *Israel J. Med. Sci.* 5:772, 1969.
315. Rodak, J., National Center for Health Care Technology, Rockville, Md., personal communication, December 1980.
316. Roe, B. B., "The UCR Boondoggle: A Death Knell for Private Practice?" *N. Eng. J. Med.* 305(1):41, 1981.
317. Rogers, E. M., and Shoemaker, F. F., *Communication of Innovations: A Cross-Cultural Approach,* 2d ed. (New York: Free Press, 1971).
318. Romm, F. J., and Hulka, B. S., "Developing Criteria for Quality of Care Assessment: Effect of the Delphi Technique," *Health Serv. Res.* 14:309, 1979.
319. Roos, L. L., Jr., et al., "Assessing the Impact of Tonsillectomies," *Med. Care* 16(6):502, 1978.
320. Roos, L. L., Jr., et al., "Using Administrative Data Banks for Research and Evaluation: A Case Study," *Evaluat. Q.* 3(2):236, 1979.
321. Roos, N. P., "Who Should Do the Surgery? Tonsillectomy-Adenoidectomy in One Canadian Province," *Inquiry* 16(1):73, 1979.
322. Roos, N. P., et al., "A New Audit Procedure Applied to an Old Question: Is the Frequency of T&A Justified?" *Med. Care* 15(1):1, 1977.
323. Rosati, R., et al., *Expansion, Extension, and Evaluation of a Clinical Databank for Coronary Artery Disease,* contract No. HRA 230-76-0300, submitted to the National Center for Health Services Research, Hyattsville, Md., August 1980.
324. Rosenthal, R., "Combining Results of Independent Studies," *Psychol. Bull.* 85:185, 1978.
325. Rosenthal, R., *Experimenter Effects in Behavioral Research* (New York: Irvington, 1976).
326. Ross, R. S., "Ischemic Heart Disease: An Overview," *Am. J. Cardiol.* 38:496, 1975.
327. Rothwell, R., and Zegveld W., "The Role of Small Manufacturing Enterprises in Innovation," *P. Innov.* 2(1):3, 1979.
328. Rubenstein, L., et al., "Quality of Care Assessment by Process and Outcome Scoring," *Ann. Intern. Med.* 86:617, 1977.
329. Ruffin, J. M., et al., "A Co-Operative Double-Blind Evaluation of Gastric Freezing in the Treatment of Duodenal Ulcer," *N. Eng. J. Med.* 281:16, 1969.
330. Runyon, R. P., and Haber, A., *Fundamentals of Behavioral Statistics,* 3d ed. (Reading, Mass.: Addison-Wesley, 1976).
331. Russell, L. B., *Technology in Hospitals: Medical Advances and Their Diffusion* (Washington, D.C.: The Brookings Institution, 1979).
332. Ryder, H. W., letter to the editor, *N. Eng. J. Med.* 296:45, 1977.
333. Salkever, D., and Bice, T., *Hospital Certificate-of-Need Controls: Impact on Investment, Costs, and Use* (Washington, D.C.: American Enterprise Institute for Public Policy Research, 1979).
334. Salkever, D., and Bice, T., "The Impact of Certificate-of-Need Controls on Hospital Investment," *Milbank Mem. Fund Q.* 54:185, 1976.
335. Saxe, L., and Fine, M., *Social Experiments: Methods for Design and Evaluation* (Beverly Hills, Calif.: Sage Publications, 1981).
336. Schoenbaum, S. C., et al., "The Swine-Influenza Decision," *N. Eng. J. Med.* 295:759, 1976.
337. Schor, S., and Karten, I., "Statistical Evaluation

of Medical Journal Manuscripts," *J.A.M.A.* 195:145, 1966.

338. Schulz, S. C., et al., "Dialysis in Schizophrenia: A Double-Blind Evaluation," *Science* 211:1066, 1981.
339. Scriven, M., "The Methodology of Evaluation," in *Perspectives of Curriculum Evaluation,* R. W. Tyler, et al. (eds.) (Chicago: Rand McNally, 1967).
340. Sechrest, L., et al., "Introduction—Some Neglected Problems in Evaluation Research: Strength and Integrity of Treatments" in *Evaluation Studies Review Annual,* vol. 4, L. Sechrest, et al. (eds.) (Beverly Hills, Calif.: Sage Publications, 1979).
341. Shapiro, S., "Evidence on Screening for Breast Cancer From a Randomized Trial," *Cancer* 39:2772, 1977.
342. Sherman, M. A., "Mobile Intensive Care Units: An Evaluation of Effectiveness," *J.A.M.A.* 241:1899, 1979.
343. Sherman, M. A., et al., "Threats to the Validity of Emergency Medical Services Evaluation," *Med. Care* 17:127, 1979.
344. Shields, S., Hospital Utilization Plan, Pittsburgh, Pa., personal communication, May 8, 1981.
345. Shrader-Frechette, A., "Technology Assessment as Applied Philosophy of Science," *Sci. Technol. Human Values* 6:33, 1980.
346. Silverman, W. A., "The Lesson of Retrolental Fibroplasia," *Sci. Am.* 236(6):100, 1977.
347. Silvers, J. B., "Competitive Economics of Health Care Providers in the 1980's: The Case of the Large National Clinics," presented at the Project HOPE conference, "The Role of Major Medical Clinics in Meeting Future National Health Needs," Millwood, Va., Mar. 11-13, 1981.
348. Singer, D., et al., "Unexpected Readmissions to the Coronary Care Unit During Recovery From Acute Myocardial Infarction," *N. Eng. J. Med.,* 304(11):625, 1981.
349. Singer, I. D., Washington, D.C., "Notes on Federal Efforts in Ethics and Health Policy," personal communication, 1981.
350. Siperstein, M. D., et al., "Control of Blood Glucose and Diabetic Vascular Disease," *N. Eng. J. Med.* 296:1060, 1977.
351. Skolnick, M., "The Utah Geneological Data Base: A Resource for Genetic Epidemiology," in *Banbury Report 4: Cancer Incidence in Defined Populations* (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory, 1980).
352. Skolnick, M., et al., "A Computerized Family History Data Base System," *Sociol. & Soc. Res.,* 63(3):507, 1979.
353. Smith, M. L., and Glass, G. V., "Meta-Analysis of Psychotherapy Outcome Studies," *Am. Psychol.* 32:752, 1977.
354. Smith, M. L., et al., *The Benefits of Psychotherapy* (Baltimore: Johns Hopkins University Press, 1981).
355. Smithson, L., *Structure of the U.S. Medical Supply, Equipment, and Devices Industries* (Stanford, Calif.: SRI International, 1979).
356. Sonnenfeld, J. A., "The Blue Cross and Blue Shield Associations' Medical Necessity Program," *Volunt. Effort Q.* 3(2):5, 1981.
357. Speer, D. C., and Tapp, J. C., "Evaluation of Mental Health Service Effectiveness: A "Start Up" Model for Established Programs," *Am. J. Orthopsych.* 46:217, 1976.
358. Spodick, D. H., "The Surgical Mystique and the Double Standard: Controlled Trials of Medical and Surgical Therapy for Cardiac Disease," *Am. Heart J.* 85:579, 1973.
359. Stein, D. J., "Manufacturers of Medical Devices Join the Chorus of Regulatory Critics," *National J.* 12(38):1556, 1980.
360. Stinson, E. R., and Mueller, D. A., "Survey of Health Professionals' Information Habits and Needs: Conducted Through Personal Interviews," *J.A.M.A.* 243(2):140, 1980.
361. Tamborlane, W. V., et al., "Restoration of Normal Lipid and Amino Acid Metabolism in Diabetic Patients Treated With a Portable Insulin-Infusion Pump," *Lancet* 1:1258, 1979.
362. Tattersall, R. B., "Home Blood Glucose Monitoring," *Diabetologia* 16:71, 1979.
363. Temple, R., and Pledger, G. W., "The FDA's Critique of the Anturane Reinfarction Trial," *N. Eng. J. Med.* 303:1488, 1980.
364. Thacker, S., Center for Disease Control, Atlanta, Ga., personal communication, April 1981.
365. Thibault, G. E., et al., "Medical Intensive Care: Indications, Interventions, and Outcomes," *N. Eng. J. Med.,* 302(17):938, 1980.
366. Thomas, L., *Lives of a Cell* (New York: Banton Books, Inc., 1975).
367. Thompson, M. S., and Cohen, A. B., "Decision Analysis: Electronic Fetal Montoring," in *Methods of Evaluating Health Services,* P. M. Wortman (ed.) (Beverly Hills, Calif.: Sage Publications, 1981).
368. Thornell, C. A., *Comparison of Strategies for the Development of Process Measures in Emergency Medical Services,* NCHSR Research Summary Series (Hyattsville, Md.: National Center for Health Services Research, 1981).
369. Tilson, D., et al., "The Federal Interest," in *The Diffusion of Medical Technology,* G. Gordon and

G. L. Fisher (eds.) (Cambridge, Mass.: Ballinger, 1975).

370. Toffler, A., *The Third Wave* (New York: William Morrow & Co., 1980).
371. Tosteson, D. C., "Lithium and Mania," *Sci. Am.* 244:164, 1981.
372. Towery, O. B., and Perry, S., "The Scientific Basis for Coverage Decisions by Third-Party Payers, *J.A.M.A.* 245(1):59, 1981.
373. Tribe, H., "Technology Assessment and the Fourth Discontinuity: The Limits of Instrumental Rationality," *S. Calif. Law Rev.* 46:617, 1973.
374. Tversky, A., and Kahneman, D., "Judgment Under Uncertainty: Heuristics and Biases," *Science* 185:620, 1974.
375. Utterback, J. M., "Innovation in Industry and the Diffusion of Technology," *Science* 183:620, 1974.
376. Utterback, J. M., and Abernathy, W. J., "A Dynamic Model of Process and Product Innovation," *Omega* 3(6):639, 1975.
377. Veatch, R. M., Bethesda, Md., "Consensus Development in Medical Technology: Problems for Planning a Consensus Conference on Coronary Artery Bypass Surgery," unpublished, 1980.
378. Veatch, R. M., and Branson, R. (eds.), *Ethics and Health Policy* (Cambridge, Mass.: Ballinger, 1976).
379. Wagemaker, H., Jr., and Cade, R., "The Use of Hemodialysis in Chronic Schizophrenia," *Am. J. Psychiatry* 134:684, 1977.
380. Wagner, J. L., "Toward a Research Agenda on Medical Technology," in *Medical Technology,* publication No. (PHS) 19-3254 (Washington, D.C.: U.S. Government Printing Office, 1979).
381. Wangensteen, O. H., et al., "Achieving 'Physiological Gastrectomy' by Gastric Freezing," *J.A.M.A.* 180:95, 1962.
382. Warner, K. E., "A 'Desperation-Reaction' Model of Medical Diffusion," *Health Serv. Res.* 10:369, 1975.
383. Warner, K. E., and Luce, B. R., *Cost-Benefit and Cost-Effectiveness Analysis in Health Care: Principles, Practice, and Potential* (Ann Arbor, Mich.: Health Administration Press, University of Michigan, in press).
384. Warren, K. S., (ed.), *Coping With Biomedical Literature* (New York: Praeger, 1981).
385. Watkins, R., Group Health Cooperative, Seattle, Wash., personal communication, May 6, 1981.
386. Weinstein, M. C., and Fineberg, H. V., *Clinical Decision Analysis* (Philadelphia: W. B. Saunders, 1980).
387. Weiss, D. H., *Evaluation Research: Methods for Assessing Program Effectiveness* (Englewood Cliffs, N.J.: Prentice-Hall, 1972).
388. Wennberg, J. E., "Commentary: Physician Uncertainty, Specialty, Idealogy, and a Second Opinion to Tonsillectomy," *Pediatrics* 59:952, 1977.
389. Wennberg, J. E., et al., "The Need for Assessing the Outcome of Common Medical Practices," *Ann. Rev. Public Health* 1:277, 1980.
390. Wennberg, J. E., et al., "Some Uses of Claims Data for the Analysis of Surgical Practices," in *Proceedings of the 18th National Public Health Conference on Records and Statistics,* DHHS publication No. (PHS) 81-1214, August 1980.
391. Whittaker, G. F., "The Evaluation of a Merged Hospital System: Application of the Interrupted Time-Series Design," *Evaluat. Rev.* 5:68, 1981.
392. Willems, J., "The Relationship Between the Diffusion of Medical Technology and the Organization and Economics of Health Care Delivery, in *Medical Technology,* J. Wagner (ed.), DHEW publication No. (PHS) 79-3254, 1979.
393. Williams, M. E., University of Illinois, Champaign-Urbana, Ill., personal communication, March 1981.
394. Williams, M. E., et al. (ed.), *Computer Readable Data Bases: A Compendium and Source Book* (Washington, D.C.: American Society for Information Science, 1979).
395. Williams, M. E., "Database and Online Statistics for 1979," *ASIS Bulletin* 7(2):27, 1980.
396. Williamson, J. W., Johns Hopkins University, "Reformation of Project Goals," unpublished paper, 1981.
397. Wolf, C., Jr., *Ethics and Policy Analysis* (Santa Monica, Calif.: Rand Corp., February 1980).
398. Wolf, S., "High Time for Synthesis," *Forum on Medicine* 3:502, 1980.
399. Worthen, D. B., "Prescribing Influences: An Overview," *Br. J. Med. Educ.* 7:109, 1973.
400. Wortman, P. M., "Consensus Development," in *Methods for Evaluating Health Services,* P. M. Wortman (ed.) (Beverly Hills, Calif.: Sage Publications, 1981).
401. Wortman, P. M., "Randomized Clinical Trials," in *Methods for Evaluating Health Services,* P. M. Wortman (ed.) (Beverly Hills, Calif.: Sage Publications , 1981).
402. Wortman, P. M., et al., "The First Year of the Education Voucher Demonstration: A Secondary Analysis of Student Achievement Test Scores," *Eval. Q.* 2:193, 1978.

403. Wylie, J. H., et al., "Effect of Cimetidine on Surgery for Duodenal Ulcer," *Lancet* 1:1307, 1981.
404. Young, D. A., "Communications Linking Clinical Research and Clinical Practice," in *Biomedical Innovation,* E. B. Roberts, et al. (eds.) (Cambridge, Mass.: MIT Press, 1981).
405. Zaltman, G., et al., *Innovations and Organizations* (New York: John Wiley & Sons, 1973).
406. Zimmerman, H., Rhode Island Health Services Research, Inc., Providence, R. I., personal communication, Apr. 17, 1981.
407. Zwick, D., Bethesda, Md., personal communication, December 1981.

Index

Index

abortion, 48
ananencephaly, 48
arthroscopy, 82
artificial heart, 48-49, 75
assessment
  clinical trials in, 130-137
  compendium of bibliographic data bases for (table), 121-126
  consensus development, 64, 100, 144-145
  cost-benefit analysis in, 11, 39
  cost-effectiveness analysis in, 11, 14, 39, 40, 41, 46, 60-62, 67
  critique of current system for, 91-102
  data sources for, 24-27
  dissemination of information on, 101-102
  of economic effects, 24
  encouragement of private sector leadership in, 108
  experimental studies in, 34-35
  group decision methods in, 62-66, 93
  of health effects, 23, 33-39
  identification of medical technologies needing, 94-97
  information for, 23-30
  interpretation of studies, 128-137
  integrated system of, 105
  literature review in, 57-58
  methodological considerations in, 127-149
  objective of, 91
  of performance standards, 23-24
  Process phases, 94-102
  proposed organization for, 106-107
  relationship to innovation, 146-149
  of social effects, 24
  social values in, 45-53
  statistical data sources for (table), 113-121
  synthesis activities in, 57-60, 100-101, 137-146
  testing phase, 98-100
  uncontrolled study designs in, 136-137
  validity of research, 128-129

Bibliographic Retrieval Service, 28
BIOSIS, 28
Blue Cross/Blue Shield, 7, 26, 81, 92
  Medical Necessity Project, 78, 98
Blue Cross of Massachusetts, 78
Blue Shield of California, 78, 82
  Ambulatory Surgery Project, 78
  Medical Policy Committee, 78
Brook, Robert, 192
Bunker, John, 192

caesarian delivery, 52, 129
California Podiatry Association, 78
cancer
  chemotherapy, 75
  endometrial, 38
  Halsted radical surgery, 129
  screening, 75
case-control study, 37-38
case studies, 10-11, 128
Centers for Disease Control (CDC), 13, 51, 85
Central Intelligence Agency (CIA), 13, 51
chemotherapy, 37, 75
clearinghouses, 29-30
Cleveland Clinic, 77
clinical trials, 131-137
  cohort design in, 35-36, 132
  historical controls, 134-135
  matching in, 132-133
  retrospective case-control studies, 133-134
  time-series design, 135-136
cobalt therapy, 83
Commission on Professional and Hospital Activities (CPHA), 26-27
Community Services Administration, 29
computed tomography (*see* CT scanning)
Congress
  acts of (*see* legislation)
  House Committee on Energy and Commerce, 193
  options for (*see* Policy options), 18
  Senate Committee on Labor and Human Resources, 193
consensus development, 63
construct validity, 129, 139
coronary artery bypass graft surgery, 52, 129, 131
coronary care units (CCUs), 37
Coronary Drug Project, 130
cost-effectiveness analysis (CEA), 61
costs vs. prices, 40-41
CT scanning, 40, 46, 78, 84

data bases, 25, 78
  Basic Vital Statistics, 113
  Bioethicsline (table), 121
  BIOSIS Previews (table), 121
  Blue Cross/Blue Shield Systems (table), 117
  Boston Collaborative Drug Surveillance Program (table), 119
  Boston University Medical Center Drug Epidemiology Unit (table), 119
  Bureau of Labor Statistics Annual Occupational Injuries and Illness Survey (table), 117
  Cambridge Scientific Abstracts (table), 122, 125
  CANCERLIT (table), 121
  CANCERPROJ (table), 121
  Centers for Disease Control Birth Defects Monitoring System (table), 114
  Centers for Disease Control Hepatic Angiosarcoma Surveillance Finding Effort (table), 114
  Centers for Disease Control National Disease Surveillance Program (table), 113
  Centers for Disease Control Smokey Nuclear Test Cohort Study (table), 114
  Chemical Abstracts Service (table), 121
  Child Abuse and Neglect (table), 122
  CLINPROT (table), 122
  compendium of bibliographic (table), 121-126

Comprehensive Dissertation Index (table), 122
Conference Papers Index (table), 122
Department of Commerce Census of the Population (table), 117
Department of Defense Civilian Health and Medical Program of the Uniformed Services (table), 117
Department of Defense TRI-Service Medical Information Systems (table), 117
DES Vaginal Cancer Registry (table), 118
DRUGINFO and Alcohol Use/Abuse (table), 122
Duke University Cardiovascular Data Bank (table), 118
Dunedin Program (table), 119
ENVIRONLINE (table), 122
EPILEPSYLINE (table), 122
Excerpta Medica EMBASE (table), 122
Foods ADLIBRA (table), 122
Food and Drug Administration Adverse Drug Reaction Spontaneous Recording System (table), 114
Food and Drug Administration Boston Collaborative Drug Surveillance Program (table), 119
Food and Drug Administration for Implanted Cardiac Pacemakers (table), 115
Food and Drug Administration Medically Oriented Data System (table), 115
Food and Drug Administration Nationwide Evaluation of X-Ray Trends (table), 115
Food and Drug Administration National Registry for Drug-Induced Ocular Side Effects (table), 115
Food and Drug Administration Poisoning Control Case Registry (table), 114
Food and Drug Administration Registry of Adverse Reactions to Contrast Media (table), 115
Food and Drug Administration Registry of Dermatological Reactions to Drugs (table), 114
Food and Drug Administration Registry of Tissue Reactions to Drugs (table), 114
Health Index Info (table), 122
Health Planning & Admin, 28, 29
Health Planning and Administration data base (table), 122
HISTLINE (table), 122
Hospital Discharge Data Systems (table), 118
International Pharmaceutical Abstracts (table), 122
IRL Life Sciences Collection (table), 122
Johns Hopkins University POPLINE (table), 122
Kaiser-Permanente Health Services Research Center (Oregon Region) System (table), 120
Laboratory Animal Data Bank (table), 122
Manitoba Health Services Commission Databank (table), 120
Massachusetts General Hospital Computer-Stored Ambulatory Record System (table), 118
Massachusetts General Hospital Intensive Care Databank (table), 119
Medicare Claims File (table), 116
Medicaid Management Information System (table), 116
MEDLINE (table), 122
MEDOC (table), 122
National Academy of Dermatology Registry of Dermatological Reactions to Drugs (table), 114
National Cancer Institute Surveillance, Epidemiology, and End Results Program (table), 115
National Clearinghouse for Mental Health Information (table), 122
National Health Interview Survey (table), 113
National Health and Nutrition Examination Survey (table), 113
National Hospital Discharge Survey (table), 113
NCHS Basic Vital Statistics (table), 113
NCHSR Computer-Stored Ambulatory Record System (table), 118
NCHSR/Duke University Cardiovascular Data Bank (table), 118
NCHS National Ambulatory Medical Care Survey (table), 113
NCHS National Death Index (table), 113
NCHS National Fetal Mortality Survey (table), 113
NCHS National Health Interview Survey (table), 113
NCHS National Health and Nutrition Examination Survey (table), 113
NCHS National Hospital Discharge Survey (table), 113
NCHS National Natality Index (table), 113
NCHSR Problem Oriented Medical Information Systems (table), 118
New York Hospital databank on patients in coma for causes other than head injury (table), 118
NHLBI Coronary Artery Surgical Study (table), 116
NHLBI Framingham Heart Study (table), 115
NHLBI Hypertension Detection Second Followup Program (table), 115
NHLBI Multirisk Factor Intervention Trial (table), 115
NHLBI Percutaneous Transluminal Coronary Angioplasty Registry (table), 116
NIAAA Alcoholism Program Monitoring System (table), 114
NIAMDD National Diabetes Data Group (table), 116
NIDA Client Oriented Data Acquisition Process (table), 114
NIDA Drug Abuse Warning Network (table), 114
NIDA National Drug Abuse Treatment Utilization Survey (table), 114
NIOSH Registry of Toxic Effects of Chemical Substances (table), 125
Olmstead County, Minnesota System (table), 119
Pediatric Drug Surveillance Program (table), 119
Pittsburgh Diabetes Registry (table), 118
PRE-MED (table), 125
PSRO Hospital Discharge Data Set (table), 116
Rhode Island Diabetes Registry (table), 118
Rhode Island Health Services Research, Inc. SEARCH (table), 119
RINGDOC (table), 125
Science Citation Index SCISEARCH (table), 125
Seattle Group Health Cooperative System (table), 119
Seattle Heart Watch (table), 120
SERLINE (table), 125
SOCIAL SCISEARCH data base (table), 125
SSI Annual Disability Determination/Social Security Income Extract (table), 117

  SSI Continuous Disability History Survey (table), 117
  SSIE (table), 126
  SSI Leed File (table), 117
  statistical (table), 113-120
  Toxicology Data Bank (table), 126
  TOXLINE (table), 126
  University of North Carolina Population Bibliography (table), 124
  University of Vermont Problem Oriented Medical Information Systems (table), 118
  Utah Genealogical Data Base (table), 116
  VETDOC (table), 126
  Veterans Administration Compensation and Pension System (table), 117
  Veterans Administration Patient Treatment File (table), 117
decisionmaking, 66-67
Delphi technic[illegible] 15, 65, 66, 93, 145-146
  features of, 62
Department of Agriculture, 29
Department of Commerce, 29, 30
Department of Defense (DOD), 13, 30, 51, 105
Department of Education, 29
Department of Energy (DOE), 29, 30
Department of Health, Education, and Welfare (*see* Department of Health and Human Services)
Department of Health and Human Services (DHHS), 9, 16, 18, 25, 29, 80, 85, 96, 108, 109
  Ethics Advisory Board, 13, 50, 51
  Technology Coordinating Committee, 9, 84, 107
Department of Housing and Urban Development, 29
Department of Justice, 29
Department of Labor, 29
Department of Transportation (DOT), 29
DHEW (*see* Department of Health and Human Services)
DHHS (*see* Department of Health and Human Services)
DIALOG Information Service, Inc., 28
discounting, 41
DOE (*see* Department of Energy)
DOT (*see* Department of Transportation)
drugs, 9, 71-72
  advertising of, 86
  adverse reactions to, 98
  Federal regulation of, 78-79
  investigational new (IND), 79
  new drug application (NDA), 9, 12, 79
  Phase I studies of, 79
  Phase II studies of, 79
  Phase III studies of, 79
  surveillance of, 36, 97-98
Duke University, 25

electroencephalography, 83
Emergency Care Research Institute (ECRI), 78
endoscopy, 82
end-stage renal disease (ESRD), 48
End-Stage Renal Disease Program, 48
ESRD (*see* end-stage renal disease)
estrogen, 38

Excerpta Medica, 28
external validity, 129, 139

FDA (*see* Food and Drug Administration)
fetal monitoring, 57, 129
Food and Drug Administration (FDA), 3, 7, 8, 9, 17, 72, 74, 78, 79, 80, 85, 87, 88, 92, 94, 95, 96, 105, 127
  regulation of drugs and devices by, 77, 99
  reporting system for adverse drug reactions, 98
  role in medical technology assessment, 12-13
  surveillance of drugs, 97-98, 98
Fowles, Jinnet, 192

gastric freezing, 4, 130, 167
glossary, 198-202
group decision methods, 93, 144-146
  advantages and disadvantages of, 65-66

Hartford Foundation, 78
HCFA (*see* Health Care Financing Administration)
Health Care Financing Administration (HCFA), 7, 9, 17, 18, 25, 48, 52, 74, 75, 84, 87, 88, 92, 94, 95, 96, 97, 98, 99, 100, 101, 105, 106, 109
  Health Standards and Quality Bureau, 77
  Office of Coverage Policy, 77
  Office of Research and Demonstrations, 13, 76, 77, 95
  proposal to allow reimbursement for experimental technologies, 107-108
  research programs of, 76-77
  role in medical technology assessment, 9
health maintenance organizations (HMOs), 80
Health Resources Administration (HRA)
  Bureau of Health Planning, 29, 30
hemodialysis and kidney transplant surgery
  case study of, 175-184

informed consent, 49
innovation
  case studies of medical, 167-184
  characteristics of successful, 151-152
  definitions of, 150
  process for drugs and devices, 152-155
  process for medical and surgical procedures, 155-157
  public accountability in, 157-165
  research on, 150-151
Institute for Health Care Evaluation, 19
  model of, 185-191
  proposal for, 107
insurance claims records, 25
internal validity, 138-139

Johns Hopkins Hospital, 78

W. K. Kellogg Foundation, 26

legislation
  Alcohol Abuse and Alcoholism Act, 80
  Community Mental Health Centers Act, 80

  creation of the National Center for Health Care Technology (NCHCT), 106
  Food, Drug, and Cosmetic Act, 79, 93
  Freedom of Information Act, 51
  Medical Device Amendments (to Food, Drug, and Cosmetic Act), 9, 79, 127
  National Health Planning and Resources Development Act, 13, 80
  Public Health Services Act, 80
  Social Security Act, 6, 80, 81
  State certificate-of-need laws, 80
Library of Congress
  National Referral Center, 29
Lohr, Kathleen, 192

macroallocation, 49
mammary artery ligation, 130
Manitoba Health Services Commission, 36
maternal serum alpha-fetoprotein (MSAFP), 48, 96
  case study of, 174-175
Mayo Clinic, 37, 77
Medicaid, 26, 48, 76, 80, 81, 83, 98
medical devices, 72
  Federal regulation of, 79-80
Medical Literature Analysis and Retrieval System (*see* MEDLARS)
medical technologies
  assessment strategies for, 3-4, 6, 92-93
  capital investment in, 161
  conceptual framework for assessing, 7
  cost-effectivness of, 52
  definition of, 3
  development, diffusion, and use of, 71-88
  economic effects of, 11-12, 14-15, 52
  effectiveness of, 52
  efficacy of, 33
  ethical issues in, 48-49, 52
  Federal role in assessing, 4, 6
  health effects of, 10-11, 14-15, 33-39
  health professionals and, 5
  identification of assessment needs, 8-10, 16-17
  industry, 5-6, 99
  information dissemination on, 15-16, 17-18, 84-85, 85-86
  innovation process for, 71-74, 150-166
  legal implications of, 52
  marginal valuation of, 39-40
  market for, 6
  medical devices, 72, 159-160
  medical and surgical procedures, 73, 99, 160-161
  opportunity cost of, 39
  pharmaceutical industry, 71-72
  public concern over, 4-5
  regulation of, 86-87, 158-161
  reimbursement policy effects on, 81-84, 161-165
  research findings on, 14
  safety of, 33, 52
  social effects of, 12, 14-15, 47-50, 52
  testing of, 10-13, 17
  value premises of assessment of, 45-47
Medicare, 4, 6, 7, 26, 52, 75, 76, 80, 81, 83
MEDLARS, 18, 28, 29
MEDLINE, 28
  subject coverage of, 101
menisectomy, 82
meta-analysis, 58-59, 128, 142-143
methotrexate, 37
MSAFP (*see* maternal serum alpha-fetoprotein)
multiple sclerosis, 78
Multiple Sclerosis Foundation, 78

National Academy of Sciences (NAS)
  Institute of Medicine, 106
National Cancer Institute (NCI), 30
  Surveillance, Epidemiology, and End Results (SEER) program, 26
National Center for Health Care Technology (NCHCT), 9, 13, 17, 18, 50, 52, 63, 75, 77, 84, 85, 87, 94, 95, 96, 98, 108, 110, 127
  Clinical Data Acquisition Plan, 99
  health care research programs of, 76
  proposal to refund, 107
  role in medical technology assessment, 9, 13
National Center for Health Services Research (NCHSR), 9, 17, 18, 25, 75, 84, 95, 97, 99-100, 105, 109
  health services research programs of, 75-76
  User Liaison Program, 15, 84
National Center for Health Statistics (NCHS), 24-25, 27
National Commission, 13, 50-51
National Health Planning Information Center, 29
National Heart and Lung Institute (*see* National Heart, Lung, and Blood Institute)
National Heart, Lung, and Blood Institute, 26, 49
National Institutes of Health (NIH), 9, 17, 18, 25, 28, 51, 57, 63, 65, 66, 74, 77, 88, 92, 95, 96, 97, 99, 101, 105, 109
  basic research funding by, 74-75
  clinical trials supported by (table), 75
  consensus development in, 144-145
  consensus development meetings of (table), 64
  Office for Medical Applications of Research (OMAR), 16, 63, 85, 94, 98, 100, 109
  role in medical technology assessment, 9, 13
National Library of Medicine (NLM), 15, 16, 28, 29, 30, 65, 109
  data bases of, 18, 28, 101, 102
  expansion of data base to include more Governmental research reports, 110
  Hepatitis Knowledge Base, 63-64, 100-101, 102
  Lister Hill National Center for Biomedical Communications, 63
National Science Foundation (NSF), 13, 29, 51
National Technical Information Service (NTIS), 16, 28, 30, 102
  data base of (table), 122
NCHSR (*see* National Center for Health Services Research)
NIH (*see* National Institutes of Health)
NLM (*see* National Library of Medicine)
nominal group technique, 62-63

non-randomized clinical trials, 127
NSF (*see* National Science Foundation)
NTIS (*see* National Technical Information Service)

observational studies, 35-38
Office of Science and Technology (*see* White House Office of Science and Technology
Office of Technology Assessment (OTA), 3, 14, 16, 19, 24, 25, 30, 33, 39, 46, 47, 57, 60, 61, 67, 94, 109, 129
  acknowledgments to contributors to this report, 195-197
  conclusions on adequacy of testing for safety and efficacy, 100
  conclusions on communication of information on medical technologies, 102
  conclusions on identification of medical technologies, 96, 97, 98
  conclusions on synthesis phase of assessment, 101
  functions of, 50
  health-related reports of, 13
  methods used in this study, 192-194
  other volumes produced by this study, 193-194
OMAR (*see* under National Institutes of Health)
osteosarcoma, 37
OTA (*see* Office of Technology Assessment)
overhead costs, 40

Pap test program, 40
peptic ulcer, 4
percutaneous transluminal coronary angioplasty (PTCA), 26, 87
  case study of, 168-174
Peter Bent Brigham Hospital, 78
phakoemulsification, 82
Phase I studies, 12, 79
Phase II studies, 12, 79
Phase III studies, 12, 79
plasmapheresis, 78
policy options, 18, 105-110
  for congressional oversight, 108-110
  for coordination of technology assessment, 110
  for dissemination activities, 109-110
  for education of researchers, 108
  for identifying medical technologies for assessment, 109
  legislative, 106-109
  organizational, 106-107
  for private sector oversight, 108
  for research funding, 107-108
  for synthesizing research information and group decisionmaking activities, 109
  for testing medical technologies, 109
policy-related activities, 50-52, 74-86
President's Commission, 13, 50-52
prices vs. costs, 40-41
Professional Standards Review Organizations (PSROs), 13, 80, 92, 95, 97, 98, 101, 105, 106, 109
Project Share, 29-30
prospective studies, 10
PSROs (*see* Professional Standards Review Organizations)
psychotherapy, 59, 73
Public Health Service, 25, 77, 96, 107, 109
  Health Services Research Coordinating Committee, 84

radiology, 52, 83
Rand Corporation, 62
randomized clinical trials (RCT), 10, 11, 34-35, 39, 57, 74, 78, 91, 93, 105, 106, 109, 127, 128
  attributes of, 130-131
  basis for conducting, 149
RCT (*see* randomized clinical trials)
Reiss, John, 192
research and development (R&D), 11, 15
  costs of, 40
  in development, diffusion, and use of medical technologies, 74-78
  Federal activities in, 75-77
  private sector activities in, 77-78
  protection of human subjects, 50-52
research (*see* research and development)
Royal Perth Hospital, 37

schizophrenia
  hemodialysis treatment of, 167-168
SEER program (*see* under National Cancer Institute)
Simpson's paradox, 58
Southwestern Michigan Hospital Council, 27
spina bifida, 48
Stanford University, 25
statistical conclusion validity, 34, 129, 139
synovectomy, 82
System Development Corp., 28

third-party payers, 73, 81-84, 86-87
tomography (*see* CT scanning)
Toxicology Information Online (*see* TOXLINE)
TOXLINE, 28

Veterans Administration, 13, 51, 105, 130
vital statistics, 25
voting method, 58, 141

Wennberg, John, 192
White House Office of Science and Technology, 13, 51
Williamson, John, 192
Winkler, John, 192
Woolf, Patricia, 192
Wortman, Paul, 192

Office of Technology Assessment

The Office of Technology Assessment (OTA) was created in 1972 as an analytical arm of Congress. OTA's basic function is to help legislative policymakers anticipate and plan for the consequences of technological changes and to examine the many ways, expected and unexpected, in which technology affects people's lives. The assessment of technology calls for exploration of the physical, biological, economic, social, and political impacts which can result from applications of scientific knowledge. OTA provides Congress with independent and timely information about the potential effects—both beneficial and harmful—of technological applications.

Requests for studies are made by chairmen of standing committees of the House of Representatives or Senate; by the Technology Assessment Board, the governing body of OTA; or by the Director of OTA in consultation with the Board.

The Technology Assessment Board is composed of six members of the House, six members of the Senate, and the OTA Director, who is a nonvoting member.

OTA currently has studies underway in 11 general areas: energy, international security and commerce, materials, food and renewable resources, health, biological applications, communication and information technologies, oceans and environment, and space, transportation, and innovation.